

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE MEDICINA
Departamento de Medicina



TESIS DOCTORAL

**Modelos matemáticos de pronóstico para cáncer de cérvix
localmente avanzado.**

Validación externa de los modelos existentes

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

David Lora Pablos

Directores

**Agustín Gómez de la Cámara
José Fermín Pérez Regadera Gómez
Rafael Enríquez de Salamanca Lorente**

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MEDICINA

Departamento de Medicina



Modelos matemáticos de pronóstico para cáncer de cérvix localmente avanzado.

Validación externa de los modelos existentes.

Autor:

David Lora Pablos.

Directores de tesis:

Agustín Gómez de la Cámara.

José Fermín Pérez Regadera Gómez.

Rafael Enríquez de Salamanca Lorente.

Madrid, 2017.

A Juan y Paqui.

A Rosa.

A Miguel y Berta.

Agradecimientos:

A mis compañeros de la Unidad de Investigación del Hospital Universitario 12 de Octubre, en especial al Dr. José Carlos Martínez Ávila por su dominio del alemán, a Clara Calleja Fernández por su inglés, a Cristina Martín Arriscado Arroba por su ayuda en la revisión, a Mariano Barroso García por mantenerme actualizado y en red, a Pilar Cancelas Navia por su entusiasmo, al Dr. Miguel Menéndez Orenga por su estilo y a Rosa María Vega Viaña por darme las claves de la buena práctica clínica.

Mostrar mi gratitud con José Manuel Estrada Lorenzo y las personas que componen la Biblioteca del Hospital 12 de Octubre, cuya diligencia y favor han sido determinantes en el desarrollo de la presente Tesis Doctoral.

Agradecer sinceramente al Dr. Joaquín Arenas Barbero, su cercanía y atención.

Al Dr. Rafael Enríquez de Salamanca Lorente, cuyo amparo y tino han encaminado desde el comienzo a buen fin la intención de la Tesis Doctoral. Gracias por ser mi profesor.

Al Dr. José Fermín Pérez Regadera Gómez, sus clases y explicaciones sobre medicina en el ámbito de la oncología radioterápica nutrieron mis ideas. Su apoyo y confianza fueron el pilar de esta obra. Gracias por ser mi maestro.

Al Dr. Agustín Gómez de la Cámara, médico amable, generoso en la enseñanza, de ideas y frases brillantes que ilustran y fortalecen mis horas de estudio. Gracias por cultivar en mí el deseo de conocer y comprender.

Índice de Figuras.....	5
Índice de Tablas.....	6
Índice de Ecuaciones.....	8
Abreviaturas y Símbolos.....	9
1 RESUMEN.....	11
1.1 Título.....	11
1.2 Antecedentes.....	11
1.3 Objetivos.....	11
1.4 Métodos. Búsqueda y extracción de datos.....	11
1.5 Pacientes de estudio.....	12
1.6 Análisis de datos.....	12
1.7 Resultados.....	13
1.8 Conclusión.....	14
2 ABSTRACT.....	15
2.1 Title.....	15
2.2 Background.....	15
2.3 Objectives.....	15
2.4 Search methods and data collection.....	15
2.5 Patients.....	16
2.6 Data analysis.....	16
2.7 Results.....	17

2.8	Conclusion.	18
3	INTRODUCCIÓN.	19
3.1	Epidemiología del cáncer de cérvix.	19
3.2	Historia natural del cáncer de cérvix.	19
3.3	Diagnóstico y patología.	22
3.4	Estadio y evolución de la enfermedad. Estadios FIGO.	23
3.5	Tratamiento del cáncer de cérvix.	25
3.6	Modelos de pronóstico en el cáncer de cérvix.	27
3.7	Utilidad clínica de los modelos de pronóstico en el cáncer de cérvix.	32
3.7.1	Análisis de utilidad.	33
3.7.2	Análisis de la curva de decisión.	36
3.8	Finalidad de la tesis doctoral.	41
4	HIPÓTESIS.	43
5	OBJETIVOS.	45
5.1	Objetivos principales.	45
5.2	Objetivos secundarios.	45
6	MATERIAL Y MÉTODOS.	47
6.1	Revisión sistemática.	47
6.1.1	Métodos de la revisión sistemática.	47
6.2	Cohorte de pacientes para la validación externa.	51
6.2.1	Diagnóstico y Tratamiento.	52

6.2.2	Seguimiento de las pacientes.	54
6.2.3	Tratamiento de la información.	54
6.2.4	Validación externa.	55
6.3	Estructura y síntesis del documento de la tesis doctoral.	67
7	RESULTADOS.	69
7.1	Descripción de los estudios.	69
7.1.1	Resultados de la búsqueda.	69
7.1.2	Estudios incluidos.	70
7.2	Validación externa.	87
7.2.1	Descripción de la muestra de validación.	87
7.2.2	Descripción de las muestras de desarrollo.	88
7.2.3	Validación externa de los modelos de pronóstico.	93
7.2.4	Análisis de las curvas de decisión.	103
8	DISCUSIÓN.	109
8.1	Resumen de los principales resultados.	109
8.2	Exhaustividad y aplicabilidad de la evidencia.	110
8.3	Calidad de la evidencia.	114
8.4	Sesgos potenciales en el proceso.	114
8.5	Acuerdos y desacuerdos con otros estudios o revisiones.	121
8.6	Implicaciones para la práctica.	125
8.7	Implicaciones para la investigación.	125

9	CONCLUSIONES.....	127
10	BIBLIOGRAFÍA.....	129
11	ANEXOS.....	141
11.1	Estrategias de búsqueda.....	141
11.1.1	Estrategia de búsqueda en Medline (pubmed):.....	141
11.1.2	Estrategia de búsqueda en ISI Web Of Knowledge:.....	143
11.1.3	Estrategia de búsqueda en EMBASE (Ovid):.....	144
11.1.4	Estrategia de búsqueda en IME (Índice Médico Español):.....	145
11.1.5	Estrategia de búsqueda en Epistemonikos:.....	146
11.1.6	Estrategia de búsqueda en IBECs (isciii):.....	146
11.1.7	Estrategia de búsqueda en LILACS:.....	147
11.2	Informe del CEIC del Hospital Universitario 12 de Octubre.....	149
11.3	Características de los estudios incluidos en la revisión.....	153
11.4	Guía TRIPOD.....	163
12	PUBLICACIÓN.....	165

Índice de Figuras.

Figura 1. Modelo conceptual del cáncer de cérvix.	20
Figura 2. Árbol de decisión.....	34
Figura 3. Árbol de decisión.....	40
Figura 4. Sujetos y Eventos bajo estudio.	52
Figura 5. Análisis de la Curva de Decisión.....	64
Figura 6. Diagrama de flujo de la revisión sistemática de la literatura.....	69
Figura 7. Valoración correcto para cada ítem del TRIPOD.....	85
Figura 8. Diagrama de Venn de la recurrencia de las enfermas.	87
Figura 9. Curva de supervivencia de Kaplan-Meier para la muestra de validación.	88
Figura 10. Validación externa del modelo de supervivencia global de PG Rose (2015).	98
Figura 11. Validación externa de los modelos de S Polterauer (2012) y S Kang (2012).	100
Figura 12. Validación externa del modelo de SLE a 2 años de PG Rose (2012)	102
Figura 13. Umbrales de decisión para distintos modelos de pronóstico.....	104

Índice de Tablas.

Tabla 1. Estadificación del cáncer de cuello uterino (Estadio FIGO, 2009).	24
Tabla 2. Tabla de contingencia pronóstica.	37
Tabla 3. Tabla de clasificación. Estrategia 1.	39
Tabla 4. Tabla de clasificación. Estrategia 2.	39
Tabla 5. Calculadora online.	57
Tabla 6. Artículos seleccionados en el proceso de revisión sistemática.....	71
Tabla 7. Población, sujetos a estudio y supervivencia de los estudios.	75
Tabla 8. Modelos de pronóstico para cáncer de cérvix localmente avanzado.	79
Tabla 9. Variables de los modelos de pronóstico para supervivencia global.	81
Tabla 10. Variables de los modelos de pronóstico para supervivencia causa específica.	82
Tabla 11. Variables de los modelos de pronóstico para supervivencia libre de enfermedad. .	83
Tabla 12. Variables de los modelos de pronóstico para supervivencia específica	84
Tabla 13. Características de las mujeres de los estudios incluidos.....	91
Tabla 14. Validación externa de los modelos de pronóstico.	95
Tabla 15. Regiones de validez clínica.....	105
Tabla 16. Estrategia de búsqueda en Medline (pubmed).....	142
Tabla 17. Estrategia de búsqueda en ISI Web Of Knowledge.....	143
Tabla 18. Estrategia de búsqueda en EMBASE (Ovid).....	145
Tabla 19. Estrategia de búsqueda en IME (Índice Médico Español).....	145
Tabla 20. Estrategia de búsqueda en Epistemonikos.	146
Tabla 21. Estrategia de búsqueda en IBECs (isciii).	146
Tabla 22. Estrategia de búsqueda en LILACS.....	147
Tabla 23. Características del estudio X Li (2016)(141).....	153

Tabla 24. Características del estudio SH Shim (2013) (119).....	154
Tabla 25. Características del estudio S Kang (2012)(41).	155
Tabla 26. Características del estudio JY Tseng (2010) (118).....	156
Tabla 27. Características del estudio JA Liang (2008)(122).	156
Tabla 33. Características del estudio S Polterauer (2012)(117).	157
Tabla 34. Características del estudio EA Kidd (2012)(120).....	158
Tabla 35. Características del estudio PG Rose (2015)(121).....	161

Índice de Ecuaciones.

Ecuación 1. Modelo de pronóstico.....	28
Ecuación 2. Magnitud del efecto	29
Ecuación 3. Beneficio por tratar	34
Ecuación 4. Riesgo por tratar.....	35
Ecuación 5. Valor esperado por tratar.....	35
Ecuación 6. Valor esperado por no tratar.....	35
Ecuación 7. Equivalencia entre el valor esperado por tratar y no tratar	35
Ecuación 8. Probabilidad de indiferencia	36
Ecuación 9. Probabilidad de indiferencia en función del beneficio y riesgo por tratar	36
Ecuación 10. Odds en contra del evento en función del beneficio y riesgo por tratar.....	36
Ecuación 11. Beneficio neto desarrollado por CS Peirce (1884).....	38
Ecuación 12. Riesgo por tratar en función del odds a favor del evento.....	38
Ecuación 13. Beneficio neto desarrollado por AJ Vickers (2006).....	38
Ecuación 14. Estadístico C	60
Ecuación 15. Fórmula interna del estadístico C.....	60
Ecuación 16. Modelo logístico de recalibración.....	61
Ecuación 17. Puntuación de Brier.....	62
Ecuación 18. R^2 de Nagelkerke	62
Ecuación 19. Umbral de relevancia. Punto A.....	65
Ecuación 20. Umbral de relevancia. Punto B	65
Ecuación 21. Umbral de relevancia. Punto C	66

Abreviaturas y Símbolos.

CEIC: Comité Ético de Investigación Clínica.

CCLA: Cáncer de Cérvix Localmente Avanzado.

ECOG: Eastern Cooperative Oncology Group.

FIGO: Federación Internacional de Ginecología y Obstetricia.

IC: Intervalo de Confianza.

IME: Índice Médico Español.

LILACS: Literatura Latinoamericana de Información en Ciencias de la Salud.

MeSH: Medical Subject Headings.

MRI: Imagen por Resonancia Magnética.

PET: Tomografía por emisión de positrones (Positron Emission Tomography).

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

RM: Resonancia Magnética.

ROC: Receiver Operating Characteristic.

RS: Revisión Sistemática.

SCC-Ag: Niveles de antígeno de células escamosas.

SUVmax: Valor máximo estandarizado de captación del tumor de cérvix.

TC: Tomografía Computarizada.

TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.

VPH: Virus del Papiloma Humano.

VIH: Virus de la Inmunodeficiencia Humana.

1 RESUMEN.

1.1 Título.

Modelos matemáticos de pronóstico para cáncer de cérvix localmente avanzado. Validación externa de los modelos existentes.

1.2 Antecedentes.

Los modelos de pronóstico son procedimientos matemáticos capaces de obtener la probabilidad de riesgo absoluta para un desenlace específico de un individuo, basado en las características biográficas, biológicas, terapéuticas y clínicas. Implican el uso de una metodología singular y compleja. Un paso decisivo para su utilización beneficiosa es realizar una correcta validación. En la actualidad, existen múltiples modelos de pronóstico desarrollados para el cáncer de cérvix localmente avanzado (CCLA), estadio IB2 a IVA de la Federación Internacional de Ginecología y Obstetricia (FIGO). No todos los modelos de pronóstico están validados externamente, condición necesaria para su utilización clínica.

1.3 Objetivos.

Validar externamente los modelos de pronóstico para el CCLA desarrollados en pacientes tratadas con quimio-radioterapia concurrente en una cohorte independiente.

Identificar y contrastar criterios de buen uso de los procedimientos de investigación pronóstica y validación de modelos predictivos.

1.4 Métodos. Búsqueda y extracción de datos.

Se realizó una Revisión Sistemática (RS) de la literatura científica de acuerdo a los criterios y procesos Cochrane para identificar los modelos desarrollados con el objetivo pronóstico de

cuantificar la probabilidad absoluta de que ocurra un desenlace (mortalidad, mortalidad por cáncer, recurrencia o fracaso específico) tiempo dependiente en mujeres diagnosticadas con CCLA y tratadas con quimio-radioterapia concurrente.

Las bases de datos consultadas fueron Medline (pubmed), ISI Web Of Knowledge y EMBASE (Ovid) hasta octubre del 2016, e Índice Médico Español (IME), Epistemonikos, IBECS (isciii) y Literatura Latinoamericana de Información en Ciencias de la Salud (LILACS) hasta junio del 2015. Un único autor revisó los artículos según título. Las citas identificadas fueron evaluadas por dos revisores de manera independiente e íntegra (título, resumen y artículo), posteriormente, confrontadas y conjuntamente seleccionadas. La extracción de los datos fue realizada por un único autor. La evaluación metodológica de los artículos se realizó con el protocolo TRIPOD (“Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis”).

1.5 Pacientes de estudio.

La validación externa se realizó en una cohorte histórica de 297 mujeres con CCLA no operadas tratadas con quimio-radioterapia concurrente de intención radical entre julio de 1999 y septiembre de 2014 en el Servicio de Oncología Radioterápica del Hospital Universitario 12 de Octubre, Madrid, España.

1.6 Análisis de datos.

La validez estadística se cuantificó en términos de discriminación, estadístico C, calibración, coeficientes de calibración y gráficos de calibración, y medidas de rendimiento global, puntuación de Brier y R^2 de Nagelkerke.

La validez clínica fue establecida a través del análisis de la curva de decisión.

1.7 Resultados.

Se identificaron 8 artículos bajo la RS. En 3 estudios hubo otras líneas de tratamiento además de la quimio-radioterapia concurrente. Se incluyeron pacientes con cáncer de cérvix en estadios tempranos junto a pacientes con CCLA en 1 artículo. Se obtuvieron 13 modelos de pronóstico tiempo-dependientes, 2 para predecir supervivencia global, 5 supervivencia causa-específica, 3 supervivencia libre de enfermedad y 3 recurrencia específica. Los factores pronóstico habituales fueron: edad, raza, estadio FIGO, tipo histológico, grado de diferenciación, afectación del ganglio linfático y su localización, volumen del tumor, valor máximo estandarizado de captación (SUVmax) del tumor de cérvix, escala ECOG (“Eastern Cooperative Oncology Group”), niveles de antígeno de células escamosas (SCC-Ag) y tratamiento recibido. La calidad de la información mostrada en los artículos fue mejorable (guía TRIPOD).

En la cohorte de validación externa, 93 mujeres fallecieron, probabilidad de supervivencia global a 5 años de 70% (Intervalo de confianza (IC) 95%: 65-76%), 82 pacientes fallecieron por cáncer, probabilidad de supervivencia causa específica a 5 años de 71% (IC 95%: 65-76%) y 104 de las 297 mujeres presentaron enfermedad durante el seguimiento, probabilidad de supervivencia libre de enfermedad a 5 años de 64% (IC 95% 58-70%). La cohorte de validación externa fue homogénea respecto a las cohortes de desarrollo en la edad, 50 años, el tipo histológico, carcinoma de células escamosas en más del 79%, y con similar ECOG. Se encontraron diferencias estadísticamente significativas entre el estadio FIGO de las pacientes de la cohorte de validación externa, 49% con estadios FIGO de IIIA, IIIB y IVA, frente al resto de cohortes, 40% como máximo, en la afectación positiva del parametrio, la hidronefrosis, la localización de ganglios positivos y la raza, 2 estudios con población asiática.

Los siguientes modelos presentaron validez estadística y clínica: el modelo de supervivencia global a 3 años desarrollado por EA Kidd (2012) (área ROC=0.69 y umbral de utilidad clínica entre 24% y 50%), el modelo de supervivencia causa específica de S Polterauer (2012) para 1 (área ROC=0.65 y umbral entre 4 y 11%) y 3 años (área ROC=0.62 y umbral entre 12 y 36%), los modelos de supervivencia libre de enfermedad a 1 año de EA Kidd (2012) (área ROC=0.64 y umbral entre 7 y 32%), y a 2 años de PG Rose (2015) (área ROC=0.70 y umbral entre 19 y 58%), y el modelo de recurrencia a distancia de S Kang (2012) a 5 años (área ROC=0.67 y umbral entre 12% y 36%).

1.8 Conclusión.

La presente tesis identificó 5 modelos de pronóstico desarrollados en mujeres con CCLA y tratadas con quimio-radioterapia con validez externa estadística y clínica: el modelo de supervivencia global a 3 años desarrollado por EA Kidd (2012), el modelo de supervivencia causa específica de S Polterauer (2012) para 1 y 3 años, los modelos de supervivencia libre de enfermedad a 1 año de EA Kidd (2012), y a 2 años de PG Rose (2015) y el modelo de recurrencia a distancia de S Kang (2012) a 5 años. No se encontró evidencia suficiente para descartar la validez externa del modelo de EA Kidd (2012) para supervivencia causa específica en los dos momentos del tiempo evaluados.

Los umbrales de utilidad clínica están relacionados de manera matemática con los valores predictivos (negativo y positivo) del modelo de pronóstico, y son determinados a través de la caracterización de la probabilidad predicha del modelo de pronóstico.

2 ABSTRACT

2.1 Title.

Prognostic mathematical models for locally advanced cervical cancer. External validation of the published models.

2.2 Background.

A prognostic model is a mathematical procedure that estimates the absolute probability of risk for a specific outcome in an individual at a given time by combining different types of information (predictors) on the patient. The prognostic models involve the use of unique and complex methodology. A decisive step to be useful is a correct validation. Currently, multiple prognostic models have been developed for Locally Advanced Cervical Cancer (LACC), International Federation of Gynecology and Obstetrics (FIGO) stage between IB2 and IVA. However, few prognostic models described to date have been independently validated by other researchers and/or elsewhere, which is a necessary condition for their clinical use.

2.3 Objectives.

To externally validate the prognostic models for LACC and who were treated with concurrent chemoradiotherapy in an independent cohort.

To identify and contrast good use criteria of procedures of prognostic research and validation of predictive models.

2.4 Search methods and data collection.

A systematic review of the literature was performed to identify publications in which prognostic models were developed to quantify the probability of a time-dependent outcome

(mortality, cause-specific mortality, disease-free survival or loco-regional failure) in women with LACC and treated with concurrent chemoradiotherapy.

The following electronic databases were searched: Medline (pubmed), ISI Web Of Knowledge and EMBASE (Ovid) until October 2016, and IME, Epistemonikos, IBECS (isciii) and LILACS until June 2015. One author reviewed potential studies for inclusion. All relevant articles found were evaluated independently by two reviewers. Any discrepancies in decisions between reviewers were resolved by consensus. For each studies data extraction were realized by one author. The guideline of Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) was employed to quantify the overall quality of reporting.

2.5 Patients.

A historical cohort of 297 women with LACC who had not undergone surgery and were treated with radical concurrent chemotherapy from 1999 to 2014 at the Department of Radiation Oncology of the Hospital Universitario12 de Octubre (H12O), Madrid, Spain, were included in the study.

2.6 Data analysis.

The statistical validity was evaluated based on discrimination, c index, calibration, calibration-in-the-large and slope, calibration plot, and overall performance measures, Brier's score and R^2 of Nagelkerke.

The clinical usefulness was estimated using the decision curve analysis.

2.7 Results.

The systematic review identified 8 studies with 13 prognostic models to predict time-dependent events (2 predict overall survival, 5 predict cause-specific survival, 3 predict disease free survival, 2 predict distant recurrence and 1 predicts para-aortic recurrence). The prognostic factors used by the identified predictive models are the age, race, FIGO stage, histological type, degree of differentiation, lymph node involvement and location, tumor volume, SUVmax of the cervical tumor, ECOG, SCC-Ag levels and treatment received. The overall quality of studies reporting was suboptimal (TRIPOD guideline).

During follow-up, 93 of 297 died, corresponding to a 5-year OS of 70% (95% CI 65-76%); 82 patients died due to disease, resulting in a CSS of 71% (95% CI: 65-76%); and 104 women experienced disease recurrence during follow-up, which equates to a 5-year DFS of 64% (95% CI: 58-70%). The H12O cohort was homogeneous compared to women from other cohorts in terms of age (average age of 50), histological type (over 79% were squamous cell) and ECOG frequency. However, the FIGO stages of patients from the validation cohort (49% of patients with FIGO IIIA, IIIB and IVA staging) significantly differed from those of other cohorts (40% maximum) (p -value < 0.05). Furthermore, the involvement of the parametrium, hydronephrosis and the locations of positive nodes also differed between cohorts. The race results depended on where the study was conducted.

The following models exhibited good external validity: the area under the receiver operating characteristic curve (AUROC) for the overall survival model at 3 years from EA Kidd (2012) was 0.69 and the threshold of clinical utility (TCU) ranged from 24% to 50%, the model of cause-specific survival at 1 (AUROC = 0.65, TCU between 4% and 11%) and 3 years (AUROC = 0.62, TCU between 12% and 36%) from S Polterauer (2012), the models of

disease free survival at 1 year from EA Kidd (2012) (AUROC = 0.64, TCU between 7% and 32%) and 2 years from PG Rose (2015) (AUROC = 0.70, TCU between 19% and 58%) and the distant recurrence model at 5 years from S Kang (2012) (AUROC = 0.67, TCU between 12% and 36%).

2.8 Conclusion.

The external validation revealed the statistical and clinical usefulness of five prognostic models published in the scientific literature: the overall survival model from EA Kidd (2012), the model of cause-specific survival from S Polterauer (2012), the models of disease free survival from EA Kidd (2012) and from PG Rose (2015) and the distant recurrence model from S Kang (2012). Insufficient evidence to reject the external validation of the model of cause-specific survival from EA Kidd (2012) was found.

The thresholds of clinical utility are mathematically related with the positive and negative predictive values of prognostic model, and can be estimated using the features of predicted probabilities of prognostic model.

3 INTRODUCCIÓN.

3.1 Epidemiología del cáncer de cérvix.

El cáncer de cérvix es una neoplasia maligna que surge en el cuello del útero, estando entre los cinco cánceres más frecuente en todo el mundo en la mujer (1,2), junto al cáncer de mama (incidencia de 43.1 casos por 100000 mujeres), colorrectal (14.3 casos por 100000 mujeres), pulmón (13.6 casos por 100000 mujeres) y de estómago (7.5 casos por 100000 mujeres). El curso del cáncer de cérvix en cada país está asociado con un conjunto común de factores de riesgo con impacto diferente. Esto es debido a los diferentes programas de prevención, detección y diagnóstico, así como el acceso y tratamiento de la enfermedad. Así, en 2012, se diagnosticaron 527624 nuevos casos de cáncer de cérvix en el mundo (2), con una incidencia estimada de 14 casos por 100000 mujeres (tasas ajustadas a la población mundial), y prevalencia de 59.6 casos por 100000 mujeres para un periodo de 5 años, siendo el responsable de 265672 muertes en el mundo (2). El 84% de los casos de cáncer de cérvix se produjeron en países subdesarrollados, países donde no existen sistemas efectivos de screening (incidencia estimada de 15.7 casos por 100000 mujeres y prevalencia de 61.1 casos por 100000 mujeres). En España en 2012, donde el cáncer de cérvix no se encuentra entre los cinco cánceres más frecuentes en la mujer, se diagnosticaron 7.8 nuevos casos por cada 100000 mujeres (prevalencia de 43.9 casos por 100000 mujeres para un periodo de 5 años), siendo la tasa de mortalidad estimada en 2.1 casos por 100000 mujeres (2).

3.2 Historia natural del cáncer de cérvix.

El cáncer de cuello uterino sobre una mujer puede ser caracterizado a través del curso esperado de la enfermedad y ser representado gráficamente, a través, de un modelo conceptual (3). La Figura 1 presenta la información organizada desde la población sana hasta

el desenlace final de la vida de la paciente diagnosticada con cáncer de cérvix. Se añaden los elementos claves en el progreso y lucha contra el cáncer de cuello uterino en los apartados de prevención (1,4,5), detección (6,7), diagnóstico (5) y tratamiento (8–10).

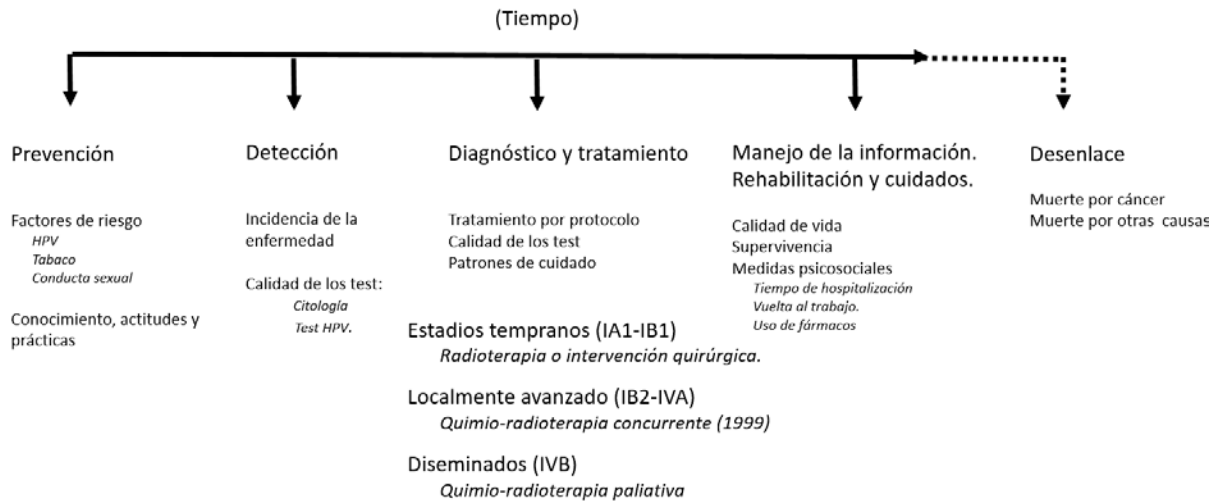


Figura 1. Modelo conceptual del cáncer de cérvix.

En la historia natural del cáncer de cuello uterino, se sabe que el cáncer de cérvix es causado en la mayoría de los casos por el virus del papiloma humano (VPH) (4,5). Bien es cierto, que no todas las infecciones por el VPH causan cáncer. Además, para que el cáncer de cérvix se desarrolle (11) influyen otros factores concurrentes con la propia infección como son (1,4): la duración de la infección, procesos de infección más largos aumenta la probabilidad de su evolución a una neoplasia cervical intraepitelial, la edad, el estado de inmunodeficiencia, (por ejemplo mujer expuesta al virus de la inmunodeficiencia humana (HIV) con test de laboratorio positivo o mujer receptora de un trasplante), el tipo de VPH, la conducta sexual (relaciones sexuales a edad temprana, menor de 16 años, o múltiples parejas sexuales, más de cuatro) y los factores reproductivos, así como, el uso de anticonceptivos orales, las infecciones por *Chlamydia trachomatis* y el hábito tabáquico, habiéndose encontrado carcinógenos específicos del tabaco e hidrocarburos aromáticos policíclicos en la mucosa del

cérvix o en el epitelio de las fumadoras (1). Otros factores importantes son el bajo estatus socioeconómicos y el déficit alimentario.

Debemos tener en cuenta que, el VPH es la infección sexual más frecuente transmitida en todo el mundo y que la mayoría de los individuos sexualmente activos de ambos sexos la adquirirán en algún momento de sus vidas (4). Se estima que la prevalencia de la infección por VPH en el cérvix es del 10-14% en la población mundial, siendo mayor en mujeres menores de 25 años, 16-19% (4). La prevalencia es diferente según localización geográfica debido a factores socioeconómicos, es decir, poblaciones sin acceso a planes de prevención o a programas de prevención más exhaustivos (por ejemplo: Australia), a factores culturales, y a las pequeñas variaciones regionales de los tipos de virus.

El contagio se produce por contacto directo de una piel a otra, o de una mucosa a otra, durante el sexo vaginal, anal u oral, estando el VPH implicado en otros tipos de cánceres, como son el cáncer de vulva, vagina, ano, pene y bucofaringe (4). Se han identificado más de 100 tipos de VPH, incluyendo 13 de alto riesgo de provocar cáncer, siendo los más frecuentes a nivel mundial: el tipo 16 (57%), 18 (16%), 58 (5%), 33 (5%), 45 (5%), 31 (4%), 52 (3%) y 35 (2%) (4).

A nivel biológico, el VPH sólo infecta a células epiteliales. El ciclo celular es modificado por el virus, prolongándolo, incluso provocando que la célula no envejezca. De esta manera, consigue la replicación viral generando miles de genomas de VPH en la propia célula. Por otro lado, la célula infectada permanece oculta al sistema inmune gracias a alteraciones genéticas provocadas por el virus que consiguen la evasión inmunológica. A pesar de todo este proceso biológico, es habitual que la mayoría de infecciones por VPH sean curadas dentro de los doce primeros meses respecto a cuando se produce la infección. El

entendimiento y comprensión de estos procesos han sido clave para el desarrollo de las vacunas, así, en la actualidad, existen vacunas monovalentes (frente al VPH 16), bivalentes (tipos 16 y 18) y tetravalentes (tipos 16, 18, 6 y 11) (7). También, se han desarrollado nuevas técnicas de cribado, consistentes en analizar muestras de células procedentes del cérvix en busca de virus de VPH (11–13).

3.3 Diagnóstico y patología.

En la población general, existen programas de cribado para la detección de lesiones precancerosas y cánceres invasivos, en fases asintomáticas o no diagnosticadas, cuestionados en determinadas sociedades para la detección del cáncer de cérvix. Además, no existen acuerdos entre las distintas sociedades médicas sobre la edad de inicio y fin de cribado, la periodicidad de la prueba y sobre la realización simultánea o no de la prueba del VPH con la citología (7). Por el contrario, las mujeres con síntomas, como sangrado vaginal anormal o dolor durante el coito son sospechosas de presentar un cáncer de cérvix. Debido al carácter transitorio e indolente de la infección por VPH, el beneficio de la detección temprana del cáncer debe equilibrarse con los daños relacionados con los procedimientos invasivos necesarios (número de pruebas invasivas, con posibilidad de sangrados vaginales, dolor e infección, afectación de la paciente debido al diagnóstico de cáncer o el propio coste de la prueba) para dar seguimiento a las pruebas de detección positivas.

La detección de un posible cáncer de cérvix a través del estudio citológico, debe ser corroborada con una biopsia. Se pueden usar diferentes tipos de biopsias: biopsia colposcópica, curetaje endocervical y biopsia de cono. Los hallazgos patológicos en la biopsia se definen histológicamente como neoplasias intraepiteliales cervicales que presentan

varios niveles de gravedad: leve, moderada y severa; carcinoma “in situ” y finalmente cáncer cervical invasivo (6).

3.4 Estadio y evolución de la enfermedad. Estadios FIGO.

Una vez diagnosticado la existencia de un carcinoma invasivo en el tejido, el paciente es clasificado según los Estadios FIGO, 2009 (5) (Tabla 1).

Estadificación de cáncer de cuello uterino (Estadios FIGO, 2009).

Estadio I	Carcinoma cervical limitado al cérvix
IA	Carcinoma invasivo pre-clínico, sólo diagnosticado microscópicamente. Invasión del estroma máximo 5 mm de profundidad desde la base del epitelio y 7 mm o menos de crecimiento horizontal
IA1	Invasión del estroma no mayor a 3 mm de profundidad y de 7 mm o menos del crecimiento horizontal.
IA2	Invasión del estroma mayor de 3 mm pero menos de 5 mm de profundidad y de 7 mm o menos de crecimiento horizontal.
<p align="center">Nota: La profundidad de la invasión de un tumor se define como la profundidad medida desde la unión epitelio-estromal hasta el lugar más profundo de la invasión. La invasión del espacio vascular venosa o linfática no afecta la clasificación. Toda lesión diagnosticada macroscópicamente incluso con invasión superficial es un IB.</p>	
IB	Lesión confinada al cérvix clínicamente visible, o lesión microscópica mayor que IA2.
IB1	Lesión clínicamente visible de 4cm o menos de dimensión máxima.
IB2	Lesión clínicamente visible de más de 4 cm de dimensión máxima
Estadio II	Carcinoma de cérvix que invade más allá del útero, pero no la pared pélvica ni el tercio inferior de la vagina.
IIA1	Lesión clínicamente visible de 4 cm o menos de dimensión máxima.
IIA2	Lesión clínicamente visible de 4 cm o más de dimensión máxima.
IIB	Tumor con invasión parametrial.
Estadio III	Carcinoma de cérvix que se extiende a la pared pélvica y/o tercio inferior de la vagina y/o produce hidronefrosis o riñón no funcionante.
IIIA	Tumor que afecta al tercio inferior de la vagina, sin extensión a la pared pélvica.
IIIB	Tumor que se extiende a la pared pélvica y/o causa hidronefrosis o riñón no funcionante
Estadio IV	Tumor que se encuentra fuera de la pelvis con infiltración de la mucosa de la vejiga o recto (IVA) o metástasis a distancia (IVB). La presencia de edema “bulloso” no constituye por sí sólo un criterio para clasificarlo como estadio IV

Tabla 1. Estadificación del cáncer de cuello uterino (Estadio FIGO, 2009).

La histopatología proporciona el diagnóstico confirmatorio de la enfermedad maligna. El estadio se determina en el momento del diagnóstico primario y nunca debería ser cambiado, aunque exista evidencia de que la enfermedad está más extendida durante la cirugía. El

estadio se determina clínicamente, basándose principalmente en el tamaño del tumor en el cuello uterino y su extensión dentro de la pelvis (1).

Diferentes estudios sobre la atención a las pacientes con cáncer de cérvix han demostrado el efecto pronóstico negativo de tener un mayor volumen, mayor extensión y mayor diseminación del tumor en la supervivencia y evolución de la enfermedad (1,14). Por tanto, el estadio clínico es un factor pronóstico fiable para estos pacientes. De esta manera, conocemos que el 100% de los pacientes con estadio IA sobrevive a 5 años, para los pacientes con estadio IB1 y pequeñas lesiones de IIA la supervivencia a 5 años disminuye a valores entre el 70-85%. La supervivencia para los tumores en estadios IB2 a IV varía en función del volumen de la enfermedad, la edad de la paciente y las comorbilidades. En general, la supervivencia libre de enfermedad es de 50-70% para los estadios IB2 y IIB, 30-50% para los estadios III, y 5-15% para los estadios IV (1).

3.5 Tratamiento del cáncer de cérvix.

El tratamiento del cáncer de cuello uterino varía en función del estadio de la enfermedad, estableciéndose diferentes grupos de tratamiento con intención curativa: cáncer de cérvix invasivo en estadios tempranos, cáncer de cérvix localmente avanzado (CCLA) y cáncer de cérvix diseminado. Este último grupo se compone de mujeres en estadios IVB y las opciones de tratamiento curativo son muy limitadas (15).

Se considera habitualmente cáncer de cérvix invasivo en estadios tempranos desde los tumores confinados al cérvix hasta aquellos con una dimensión máxima de 4 cm (estadio IB1) (16,17). Las opciones de tratamiento (17) para estas mujeres son la radioterapia o la intervención quirúrgica. La intervención quirúrgica puede variar desde la conización hasta la cirugía radical.

Se denomina CCLA a aquellos cánceres de cérvix entre el IB2 y el IVA (9). La clasificación FIGO no toma en cuenta la situación de los ganglios linfáticos, no obstante, también se considera tumores avanzados los que tienen afectación ganglionar bien pélvica o paraaórtica (16). Las pacientes con diseminación ganglionar paraaórticas son subsidiarias de tratamiento oncológico radical, ya que alrededor de la mitad pueden ser curadas a largo plazo. La terapia principal para mujeres diagnosticadas con CCLA es la quimioterapia (tratamiento farmacológico) y radioterapia (tratamiento con rayos X) concomitantes (8,10), planteada desde 1999 por el sistema de alerta del National Cancer Institute (EE.UU.) (18,19).

Idealmente, la introducción de nuevas terapias deberían ser basadas en estimaciones precisas del pronóstico basal del paciente y en las magnitudes incrementales del beneficio que puede ser obtenido, usando datos de estudios anteriores, después de evaluar el riesgo de las nuevas terapias y las opciones alternativas (20). Las revisiones sistemáticas (RS) tienen como objetivo reunir toda la evidencia empírica que cumple unos criterios de elegibilidad previamente establecidos, con el fin de responder una pregunta específica de investigación, utilizando métodos sistemáticos y explícitos, con el fin de minimizar sesgos y aportar resultados más fiables, a partir de los cuales, se puedan extraer conclusiones y tomar decisiones (21). De esta manera, se reunieron los ensayos clínicos en CCLA que comparaban la quimiorradiación concomitante con respecto al tratamiento con radioterapia y se observó el beneficio de la quimiorradiación concomitante frente a la radioterapia sola. No obstante, la quimiorradiación concomitante incrementa la toxicidad aguda con respecto al tratamiento con radioterapia, y, aún, no están claros los efectos secundarios a largo plazo de esta terapia (10). Por otra parte, datos arrojados de una RS posterior para evaluar el efecto de la quimiorradioterapia sobre todas las medidas de resultado (22) confirmó el beneficio de la quimiorradiación concomitante, su aplicabilidad a todas las mujeres, el beneficio adicional de

la quimio-radioterapia sin platino, y el posible beneficio de la quimioterapia adyuvante (tratamiento que se administra después de la terapia principal). Este último tratamiento sin pruebas suficientes que apoyen la administración de la quimioterapia adyuvante después de la quimiorradiación concomitante (23). Por tanto, aún queda por demostrar la eficacia de la quimioterapia adyuvante, evaluar sus efectos tóxicos y la calidad de vida de las pacientes (23). A través de las RS, otras estrategias de tratamiento para las mujeres con CCLA han sido evaluadas (24).

La evidencia sobre la eficacia y el pronóstico de las terapias para las mujeres con CCLA están basado en estudios observacionales y ensayos clínicos. Por tanto, la mujer diagnosticada con cáncer de cérvix y el médico se enfrentan a tomar una decisión sobre el tratamiento en un caso individual, donde el pronóstico y la respuesta al tratamiento pueden desviarse de la media del grupo (25). Identificar grupos de mujeres con características similares que realmente se beneficien de los tratamientos o sean expuestas a un menor riesgo inherente al tratamiento es conocido como medicina estratificada (25). La medicina estratificada permite orientar la terapia y tomar las mejores decisiones, pudiendo incluso, ser usada para seleccionar sujetos candidatos a ser incluidos en ensayos clínicos. De esta manera, se evita exponer a mujeres con un determinado perfil pronóstico a la nueva terapia, se crean grupos homogéneos para el evento a evaluar, y se reduce el tamaño muestral necesario de los ensayos clínicos (26–28).

3.6 Modelos de pronóstico en el cáncer de cérvix.

Se define prognosis como el conocimiento anticipado de un suceso. En medicina, se emplea para referirse al curso esperado de una enfermedad sobre un individuo, así como, a la estimación o predicción de la probabilidad de riesgo de un individuo de desarrollar un estado

de salud particular (suceso) sobre un tiempo específico, basado en su perfil clínico o no clínico (29). De esta manera, un modelo de pronóstico es un procedimiento formal que permite calcular la probabilidad absoluta de riesgo de desarrollar un evento concreto para un individuo en un instante determinado combinando diferentes tipos de información (predictores) del paciente (29,30). Esta relación, en términos de modelo lineal generalizado, puede expresarse matemáticamente como (31):

$$C(Y|X) = g(X(\beta)) \quad \text{Ecuación 1}$$

donde:

Y es la variable respuesta o dependiente.

X es la lista de variables independientes, predictoras o co-variables.

β representa los coeficientes de regresión o parámetros asociados a X.

$C(Y|X)$ es la distribución de Y dado X (como función de X).

$X\beta$ es la suma ponderada de los coeficientes o predicción lineal.

$g(.)$ es la función de enlace del modelo de regresión.

La obtención de la probabilidad de riesgo absoluta para cada individuo permite (25) identificar a los sujetos que pueden beneficiarse de las diferentes estrategias terapéuticas, conformar criterios de inclusión y exclusión de pacientes en ensayos clínicos (32), así como, ofrecer al paciente y su familia información del curso de la enfermedad, pautas a seguir y opciones de tratamiento. Los modelos multivariados, por tanto, son útiles para obtener una predicción para un sujeto (modelo de pronóstico), o establecer la magnitud del efecto (modelo explicativo) de una co-variable en la respuesta (33), donde, la magnitud del efecto de la variable j-ésima (β_j) representa el cambio del desenlace (Y), por cada unidad que se incrementa la co-variable (X_j) cuando los demás efectos permanecen constantes:

$$\beta_j = g^{-1} \left(C(Y|X_1, \dots, X_j + 1, \dots, X_n) \right) - g^{-1} \left(C(Y|X_1, \dots, X_j, \dots, X_n) \right) \quad \text{Ecuación 2}$$

La denominación de modelo de pronóstico o modelo explicativo es determinada por el objetivo por el que fue desarrollado el modelo multivariable (33). En algunas ocasiones, los modelos explicativos (34), también, son utilizados para crear criterios de inclusión en ensayos clínicos (16,35). De esta manera, bajo la asociación observada en mujeres diagnosticadas con cáncer de cérvix en estadio IB entre la recurrencia y la mortalidad con los factores de riesgo: tamaño del tumor, la profundidad en la invasión del estroma y la positividad de los ganglios linfáticos, se construyeron los criterios de inclusión (16) de un ensayo clínico (35). El ensayo clínico (35) tenía por objetivo evaluar los beneficios y riesgo de radioterapia pélvica adyuvante en la reducción de la recurrencia en pacientes con cáncer de cérvix en estadio IB. No obstante, el uso de co-variables individuales obtenidas en modelos explicativos lleva consigo una pérdida de exactitud en contraste con el uso de las probabilidades predichas a través de los modelos de pronóstico para generar criterios de inclusión en ensayos clínicos (36,37).

En lo que respecta a modelos de pronóstico, diferentes modelos predictivos han sido desarrollados en pacientes diagnosticadas de cáncer de cérvix. Ejemplo de ello son: el modelo de DY Kim (38), para mujeres con cáncer de cérvix en estadios tempranos, permite estimar la probabilidad de tener metástasis en los ganglios linfáticos con información preoperatoria, a través de una herramienta gráfica o nomograma (39,40). De ser útil (30,32), el modelo evitaría la cirugía, puesto que estas pacientes deberían ser tratadas con quimio-radioterapia y evitaría la morbilidad de tres tratamientos diferentes (cirugía, radioterapia y quimioterapia). Y otro ejemplo, el modelo desarrollado por S Kang (41) para mujeres con cáncer de CCLA para predecir la diseminación metastásica a distancia. Este modelo de pronóstico podría ser de utilidad para evaluar la supervivencia global entre la quimiorradiación concomitante

únicamente y la quimioterapia adyuvante después de la quimiorradiación concomitante (23) en mujeres con alta probabilidad de metástasis a distancia. La quimioterapia adyuvante es una terapia que intenta mejorar la supervivencia al erradicar la enfermedad metastásica subclínica residual. Esta estratificación es de interés debido a que no todas las mujeres tienen el mismo riesgo de tener una metástasis, y crearía grupos homogéneos en un ensayo clínico, donde el beneficio de la quimioterapia adyuvante podría ser atribuible a una mejoría de la supervivencia libre de enfermedad (8,10,22,42).

En un marco teórico ideal y de manera general, el desarrollo de un modelo de pronóstico consiste (31,43–45) en crear un modelo con datos provenientes de un estudio de factores predictivos bien diseñado con un número adecuado de pacientes. El modelo debería incluir toda la información pronóstica de los datos e ignorar todas las variables ruidosas, las variables correctas deberían ser seleccionadas junto a sus interacciones, en caso de existir, y determinar la forma funcional de las variables continuas sobre el evento. El modelo así construido en el marco teórico ideal no debería ser mejorado añadiendo ningún otro factor (32). En este marco, el modelo desarrollado se evaluaría en un conjunto de nuevos pacientes, proceso denominado validación externa (32,46–48), que se refiere a la capacidad del modelo para producir predicciones precisas en muestras aleatorias de poblaciones diferentes pero relacionadas y plausibles, teniendo en cuenta que pueden existir variaciones en los métodos de recogida de datos con respecto a la muestra de desarrollo. En este conjunto nuevo de datos, datos de validación, se estima el pronóstico futuro de los nuevos pacientes de forma tan precisa como la información pronóstica, incluida en el modelo, permita. En el mejor de los casos, el resultado predicho y el evento observado serán similares, validando estadísticamente (32) los resultados obtenidos en la muestra de desarrollo. En caso contrario, el modelo debería ser actualizado (43,49,50). Finalmente, el estudio de la evaluación del

impacto de la utilización del modelo ideal debería aportarse como evidencia para ser incluido en la práctica clínica real y en las guías de recomendaciones de la práctica clínica (30). Se debe tener en cuenta que, un modelo de pronóstico para ser empleado debe ser clínicamente creíble, exacto, general, es decir, ser válido en cualquier lugar, y clínicamente efectivo, en el sentido que aporte información general a los clínicos (32). Es más, un modelo de pronóstico exacto no es beneficioso, si no es generalizable o no cambia el comportamiento del médico (51).

En la actualidad, existe gran número de modelos de pronóstico desarrollados, pocos validados externamente con muestra recogida de manera consecutiva a la finalización de la recogida de la muestra de desarrollo y aún menos validados en muestras independiente, llevada a cabo por otro investigador y/o en otro lugar (46), y la mayoría de ellos sin una clara evidencia del impacto en la conducta del médico, en el evento del paciente, o en el coste de efectividad de la atención al paciente (30). Esta realidad es similar a la encontrada en los modelos de pronóstico desarrollados para el cáncer de cérvix. Si bien es cierto, que diversos trabajos han estado enfocados a la RS de modelos de pronóstico para cáncer de cuello uterino en diferentes instantes del curso de la enfermedad (3) (Figura 1), así por ejemplo, en la fase de detección, la RS realizada por I Stegeman (52) identificó tres modelos de pronóstico (53–55) para predecir carcinoma invasivo en el tejido del cuello uterino, ninguno de ellos fue validado externamente (52). En la fase de diagnóstico, se hizo una revisión de la literatura en mujeres con cáncer de cérvix invasivo en estadios tempranos con una posterior validación externa independiente (56) de los modelos encontrados (35,57–67). Así mismo, en el cáncer de cuello uterino diseminado, existe una validación externa de carácter prospectivo (15) sobre un modelo de pronóstico desarrollado (15), pero no hay ninguna RS de la literatura. En el caso de modelos de pronóstico para mujeres con CCLA, no existe ninguna RS que los reúna y los

evalúe. Además, aquellos modelos de pronóstico en mujeres con CCLA validados externamente han sido en muestras recogidas de manera consecutiva por los mismos investigadores a la finalización de la recogida de la muestra de desarrollo (46), exceptuando una comunicación a un congreso (68).

3.7 Utilidad clínica de los modelos de pronóstico en el cáncer de cérvix.

Los modelos de pronóstico son evaluados según su validez estadística, el modelo funciona correctamente en pacientes diferentes de aquellos individuos sobre los que fue desarrollado (32), sin tener en cuenta las consecuencias y validez clínica (32,69,70). Las probabilidades predichas arrojadas por los modelos de pronóstico aportan información del paciente a los médicos (32), y son de utilidad para plantear una estrategia clínica efectiva en distintos escenarios de incertidumbre (71): 1) el tratamiento, decisión sobre qué pacientes se benefician de la intervención y debería ser utilizada, en cuales es inefectiva y no debería ser aplicada, y dónde el beneficio es suficiente para algunos sujetos, pero no para todos y, por tanto, el tratamiento debería ser consensuado con el paciente (72); 2) Información personalizada (73) a las características del sujeto, que permite al médico informar al paciente y su familia sobre la evolución de la enfermedad, así como, la pertinencia de la intervención o tratamiento a realizar (25,32,70,72,74–76); 3) Ensayos clínicos, elección de grupos de riesgo en el diseño de ensayos clínicos, donde la duda se plantea en la inclusión de todos los sujetos con la enfermedad o la inclusión de sólo aquellos sujetos con la enfermedad y con mayor riesgo potencial de desarrollar el evento que se quiere prevenir y estudiar (74). 4) Comparación de modelos de pronóstico (72).

3.7.1 Análisis de utilidad.

El dilema de asignar un tratamiento adicional, ofrecer información al paciente de la evolución de su enfermedad o decidir incluir a un sujeto en un ensayo clínico es habitual dentro de la práctica clínica del médico (71,77). La incertidumbre al diagnóstico sobre la evolución del cáncer de cérvix existe, por ejemplo, muerte debida al cáncer o recidiva de la paciente. Dada esta incertidumbre, el razonamiento pronóstico del clínico (16,35–37) deseará tratar con un fármaco efectivo para el desenlace que se quiere evitar a aquellos sujetos que de no ser tratados con el nuevo fármaco sufrirán ese evento, no deseará exponer a los efectos adversos del fármaco a aquellos sujetos que de ninguna manera padecerán el desenlace de interés. La decisión sobre un sujeto basado en sus características no está exenta de error. El análisis de las decisiones clínicas (71,77,78) aborda esta problemática empleando para ello árboles de decisión, donde se describen el curso de la acción y las consecuencias de cada una de ellas, así como la probabilidad y el valor (utilidad) de cada resultado. La Figura 2 representa el marco conceptual de los árboles de decisión bajo las siguientes condiciones (71,77,78): 1) Una sola enfermedad o episodio adverso es considerada. En la Figura 2, el círculo expresa la intervención del azar. 2) Un tratamiento efectivo para la enfermedad es definido. 3) El médico debe decidir si tratar, con la incertidumbre de si el paciente va a desarrollar la enfermedad. La decisión del médico en la Figura 2 se refleja con un cuadrado. 4) Cada estrategia y desenlace está asociado con un valor de utilidad (riesgo o beneficio).

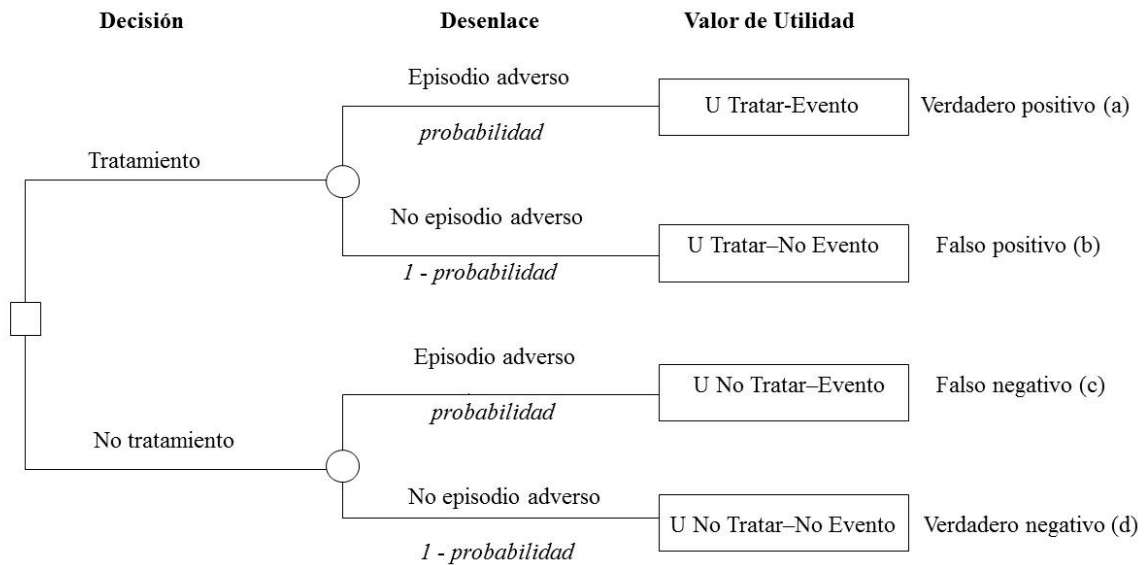


Figura 2. Árbol de decisión.

Basado en este principio, SG Pauker et al (1975) (77) derivó en el ámbito diagnóstico la relación entre el umbral de probabilidad de sospecha clínica de la enfermedad y la razón beneficio-coste (o consecuencias indeseables) de un tratamiento, que es extensible al razonamiento pronóstico (72). La clasificación tratar o no tratar (70,72,79) dentro del ámbito diagnóstico es en el ámbito pronóstico equivalente a positivo o negativo para el desenlace de la enfermedad para un tiempo concreto. El beneficio y riesgo por tratar es entendido como (72,77): 1) Beneficio por tratamiento, dado el riesgo de un resultado falso negativo comparado con un verdadero positivo. Restringido solo a los pacientes con la enfermedad, expresa la diferencia entre el valor asociado a los pacientes enfermos que fueron tratados frente a aquellos enfermos que no fueron tratados e indica la consecuencia de no aplicar un tratamiento cuando puede ser beneficioso.

$$Beneficio_{Tratar} = U_{Tratar-Evento} - U_{No Tratar-Evento} \quad \text{Ecuación 3}$$

2) Riesgo por tratamiento. Complementariamente, la diferencia entre el valor asociado a los verdaderos negativos menos los falsos positivos es la consecuencia de ser tratados

innecesariamente, o riesgo asociado con un resultado falso positivo (comparado a un resultado verdadero negativo).

$$Riesgo_{Tratar} = U_{No\ Tratar-No\ Evento} - U_{Tratar-No\ Evento} \quad \text{Ecuación 4}$$

Además, el concepto de utilidad permite estimar el valor esperado de la acción de tratar (VE_{Tratar}) y el valor esperado de no tratar ($VE_{No\ Tratar}$) como suma en cada rama de la Figura 2 de la probabilidad de cada suceso por la utilidad asociada:

$$VE_{Tratar} = p * U_{Tratar-Evento} + (1 - p) * U_{Tratar-No\ Evento} \quad \text{Ecuación 5}$$

$$VE_{No\ Tratar} = p * U_{No\ Tratar-Evento} + (1 - p) * U_{No\ Tratar-No\ Evento} \quad \text{Ecuación 6}$$

donde:

VE es el valor esperado.

p es la probabilidad de ser evento.

U es el valor de utilidad en cada rama.

Aquella acción, tratar (Ecuación 5) frente a no tratar (Ecuación 6), que tenga mayor valor esperado será la preferible. Las acciones de tratar o no tratar son equivalentes cuando el valor esperado de ambas son iguales (77):

$$VE_{Tratar} = VE_{No\ Tratar}$$

$$\begin{aligned} p * U_{Tratar-Evento} + (1 - p) * U_{Tratar-No\ Evento} &= \\ &= p * U_{No\ Tratar-Evento} + (1 - p) * U_{No\ Tratar-No\ Evento} \end{aligned} \quad \text{Ecuación 7}$$

De donde se obtiene que, el valor de p es indiferente cuando:

$$p = \text{Ecuación 8}$$

$$\frac{U_{No\ Tratar-No\ Evento} - U_{Tratar-No\ Evento}}{U_{Tratar-Evento} - U_{No\ Tratar-Evento} + U_{No\ Tratar-No\ Evento} - U_{Tratar-No\ Evento}}$$

y, por tanto, podemos concluir que: si la probabilidad de la enfermedad en un paciente es mayor que el valor de indiferencia, la acción de tratar es preferible. Mientras que si el valor de probabilidad es inferior, la opción de no tratar es la adecuada. La Ecuación 8 puede reescribirse en términos de beneficio y riesgo por tratar:

$$p = \frac{Riesgo_{Tratar}}{Riesgo_{Tratar} + Beneficio_{Tratar}} \text{Ecuación 9}$$

De igual manera, se puede reformular como la odds en contra del evento (71,72,80) y la razón entre beneficio-riesgo por tratar.

$$\frac{1-p}{p} = \frac{U_{Tratar-Evento} - U_{No\ Tratar-Evento}}{U_{No\ Tratar-No\ Evento} - U_{Tratar-No\ Evento}} \text{Ecuación 10}$$

$$\frac{1-p}{p} = \frac{Beneficio_{Tratar}}{Riesgo_{Tratar}}$$

3.7.2 Análisis de la curva de decisión.

El análisis de las curvas de decisión (72) es un método para evaluar la validez clínica de los modelos de pronóstico en el marco de la teoría de la decisión. Los modelos de pronóstico permiten estimar la probabilidad de un paciente de sufrir un desenlace en el curso de la enfermedad. La probabilidad predicha de que se produzca un evento toma valores entre 0 y 100%. La probabilidad de 100% muestra la certeza de que un paciente presente el efecto

adverso, por el contrario, probabilidad igual a 0% indica que el suceso en el paciente tiene pocas posibilidades de ocurrir. Naturalmente, un paciente determinado tendrá o no tendrá el suceso. Entre ambos extremos, existe una zona de incertidumbre. El análisis de la curva de decisión determina las regiones de las probabilidades predichas donde se optimiza la correcta clasificación de los sujetos (71,77,78). Para ello, se clasifica a un sujeto como positivo para tener el desenlace si su probabilidad predicha por el modelo de pronóstico es superior a la probabilidad sugestiva (o probabilidad a priori), entendida como fuerza en la creencia de la ocurrencia de un suceso (71). Por el contrario, se clasifica a un individuo como negativo para el evento, si su probabilidad predicha es inferior al valor umbral definido. La clasificación positivo o negativo es referida en la literatura como tratar o no tratar (70,72,79), y refleja la decisión tomada, por ejemplo, incluir en un ensayo clínico o no incluir, tratamiento adicional o estándar. El contraste de la clasificación con la evolución de la enfermedad o desenlace de interés desarrollado por el paciente establece cuatro categorías fundamentales (81) (Tabla 2):

- 1) Verdadero positivo (a): evento presente y resultado positivo;
- 2) Verdadero negativo (d): evento ausente y resultado negativo;
- 3) Falso positivo (b): enfermedad ausente y resultado positivo;
- 4) Falso negativo (c): Enfermedad presente y resultado negativo.

	Evento	No Evento
Positivo Probabilidad \geq Umbral	a	b
Negativo Probabilidad $<$ Umbral	c	d

Tabla 2. Tabla de contingencia pronóstica.

Basada en la clasificación de la Tabla 2, CS Peirce (1884) (82) formuló el beneficio neto de la siguiente manera:

$$Beneficio\ Neto_{Peirce} = Beneficio_{Tratar} * \frac{a}{N} - Riesgo_{Tratar} * \frac{b}{N} \quad Ecuación\ 11$$

siendo:

N es el número de sujetos.

a es el número de verdaderos positivos.

b es el número de falsos positivos.

Beneficio neto que fue reformulado por AJ Vickers (2006) (72), al asumir que el beneficio por tratar es igual a 1, y por tanto, el riesgo por tratar es igual a la odds a favor del evento (71,72,80), obtenido tras despejar la Ecuación 10.

$$Riesgo_{Tratar} = \frac{p}{1 - p} \quad Ecuación\ 12$$

De esta manera, el beneficio neto propuesto por AJ Vickers (2006) (72) es:

$$Beneficio\ Neto = \frac{a}{N} - \frac{b}{N} \left(\frac{p}{1 - p} \right) \quad Ecuación\ 13$$

donde:

N es el número de sujetos.

a es el número de verdaderos positivos.

b es el número de falsos positivos.

p es el umbral de probabilidad sugestiva.

Bajo la Ecuación 13 (72), simplificación del beneficio neto original (82), es posible calcular el beneficio neto sobre todo el espectro de valores de la probabilidad sugestiva (valor umbral) para la acción de clasificar a los sujetos como positivos o negativos en función del valor de probabilidad del modelo de pronóstico (positivo si la probabilidad predicha es superior a la probabilidad sugestiva y negativo en caso contrario). La acción de clasificar a los sujetos como positivos o negativos para el desenlace es comparada con dos estrategias extremas: 1) Estrategia 1, clasificar a todos los sujetos como positivos, independientemente de su probabilidad, para desarrollar la enfermedad (Tabla 3); 2) Estrategia 2, establecer a todos los pacientes como negativos para el desenlace adverso (Tabla 4). La estrategia 2 tiene beneficio neto igual a 0 para cualquier valor de probabilidad sugestiva.

	Evento	No Evento
Positivo Probabilidad \geq Umbral	a + c	b + d
Negativo Probabilidad $<$ Umbral	0	0

Tabla 3. Tabla de clasificación. Estrategia 1: Todos los sujetos son positivos para el evento.

	Evento	No Evento
Positivo Probabilidad \geq Umbral	0	0
Negativo Probabilidad $<$ Umbral	a + c	b + d

Tabla 4. Tabla de clasificación. Estrategia 2: Todos los sujetos son negativos para el evento.

Conocido que la estrategia con mayor beneficio neto es la mejor opción, la evaluación de todo el espectro de valores posibles de probabilidad permite establecer las regiones de utilidad clínica del modelo de pronóstico (Apartado 6.2.4.5.1 Umbrales de relevancia en el análisis de la curva de decisión.) con respecto a las otras dos estrategias. El árbol de decisión (71,77,78) es reflejado en la Figura 3, y representa las tres estrategias anteriores posibles en el

análisis de la curva de decisión. La estrategia 1 y 2 son casos extremos de la estrategia 3 (Tabla 2).

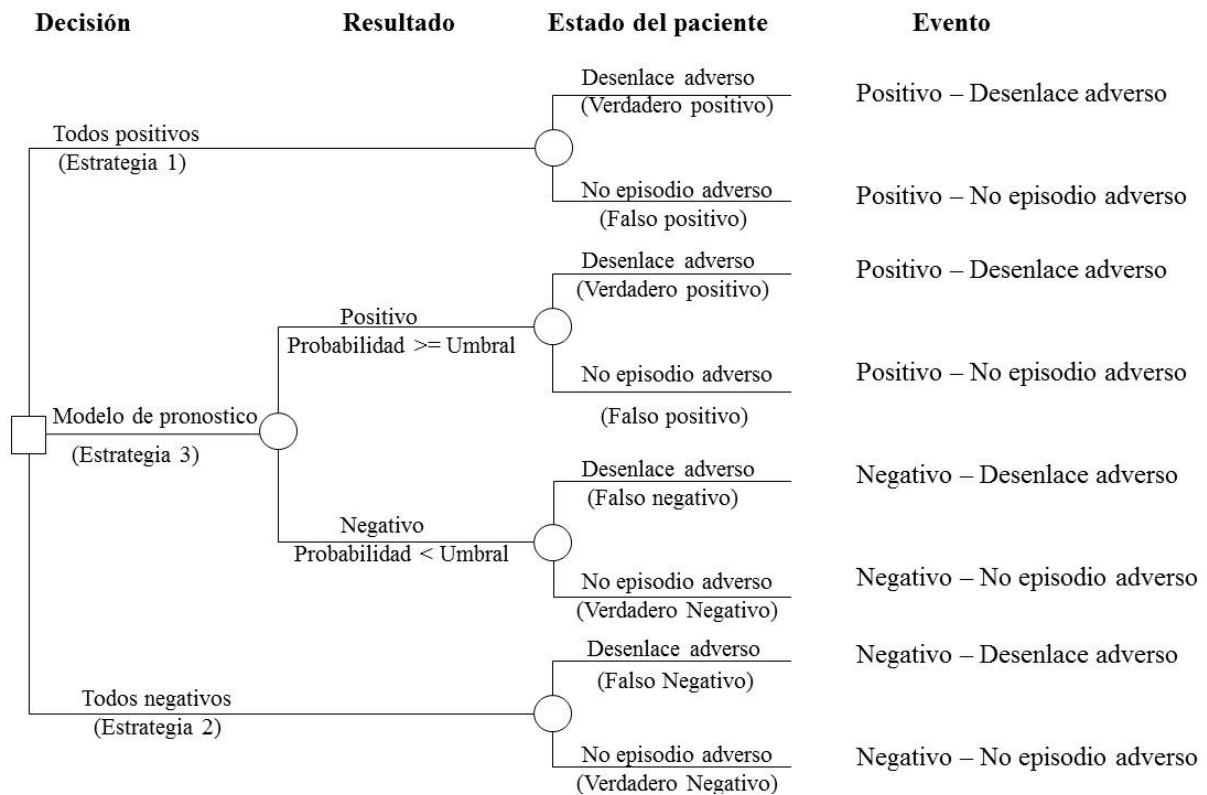


Figura 3. Árbol de decisión.

La búsqueda de los umbrales de utilidad de los modelos de pronóstico desarrollados para el CCLA nunca se ha abordado a través del análisis de las curvas de decisión (72). Aunque si se ha descrito la utilidad de puntos de corte concretos sobre las probabilidades del modelo de pronóstico para incluir o no a mujeres en ensayos clínicos, por ejemplo, en el trabajo de S Kang (2012) (41), donde se aportó para diferentes puntos de corte la repercusión que tendría en el número de pacientes estudiados, el número de pacientes no tratados innecesariamente y la población candidata. Así mismo, el artículo del modelo de DY Kim (2014) (38) para mujeres con cáncer de cérvix en estadios tempranos caracterizó en términos de sensibilidad y especificidad puntos de corte concretos de la probabilidad predicha del modelo para predecir

metástasis en los ganglios linfáticos. La dicotomización, positivo y negativo, generó dos estratos de riesgo, bajo y alto, para desarrollar metástasis en los ganglios linfáticos, de utilidad en la toma de decisión para la realización de la linfadenectomía.

3.8 Finalidad de la tesis doctoral.

En la presente tesis, se pretende validar externamente bajo los criterios de validez estadística y clínica aquellos modelos de pronóstico para predecir eventos tiempo-dependientes desarrollados para pacientes con CCLA y tratadas con quimio-radioterapia en una cohorte independiente de mujeres.

Adicionalmente, identificar y contrastar criterios de buen uso de los procedimientos de investigación pronóstica y validación de modelos predictivos.

4 HIPÓTESIS.

En la población de mujeres con CCLA y tratadas con quimio-radioterapia, es posible identificar subgrupos de pacientes con pronóstico predecible para un instante de tiempo en términos de mortalidad, mortalidad causa específica o recaída (local y/o a distancia) en el momento del diagnóstico, usando la evidencia de modelos pronósticos existentes en la literatura científica que combinan variables biográficas, biológicas, terapéuticas y clínicas.

El conocimiento de estos modelos y su aplicación sobre las mujeres diagnosticadas con CCLA permitirá estratificar a las pacientes, mejorando los resultados del tratamiento y racionalizando sus costes, la efectividad y eficiencia, mediante la aplicación de las medidas terapéuticas más pertinentes en cada subgrupo de pacientes.

5 OBJETIVOS.

5.1 Objetivos principales.

Validar externamente aquellos modelos pronósticos para predecir eventos dependientes del tiempo en mujeres con CCLA y tratadas con quimio-radioterapia en una cohorte independiente de mujeres.

5.2 Objetivos secundarios.

Analizar las características metodológicas de los trabajos con fines pronósticos publicados en la literatura médica para el CCLA.

Caracterizar las pacientes con CCLA en grupos de riesgo bajo, moderado o alto para recaída.

Determinar los valores umbrales de las probabilidades de riesgo arrojadas por los modelos predictivos que sean de utilidad clínica.

Definir y caracterizar matemáticamente la región de utilidad clínica.

6 MATERIAL Y MÉTODOS.

6.1 Revisión sistemática.

Se realizó una RS de la literatura científica para identificar aquellos artículos donde se desarrollaron modelos pronóstico para cuantificar la probabilidad de que ocurra un desenlace (mortalidad, mortalidad causa específica, supervivencia libre de enfermedad o fracaso loco-regional) tiempo dependiente en mujeres diagnosticadas con CCLA.

6.1.1 Métodos de la revisión sistemática.

6.1.1.1 Criterios para considerar estudios para esta revisión.

6.1.1.1.1 Tipos de estudios.

Los tipos de estudios considerados válidos para la revisión fueron los estudios de cohortes, estudios de casos-control, estudios de registro o estudios de tipo ensayos clínicos (43), donde se desarrollaron al menos un modelo de pronóstico y/o se crearon grupos según el riesgo, y cuyo objetivo principal fue el de cuantificar la probabilidad de que ocurra el criterio de valoración (o episodio adverso), dados las condiciones o factores incluidos en el modelo (33) (al menos una combinación de dos características o variables del individuo).

Se excluyeron los artículos de modelos pronóstico con datos ómicos, y aquellos cuyas citas bibliográficas correspondían a suplementos, comunicaciones a congreso, tesis doctorales o artículos del que no se pueda localizar el texto completo.

6.1.1.1.2 Tipos de Participantes.

Mujeres diagnosticadas con cáncer de cuello uterino localmente avanzado (7,8,22), de acuerdo a la clasificación FIGO (5): Estadio IB2, II, IIA1, IIA2, IIB, III, IIIA, IIIB y IVA, más aquellas pacientes con estadio IVB siempre que tengan linfadenectomía paraaórtica.

6.1.1.1.3 Tipos de Intervenciones.

Se seleccionan aquellos artículos donde las pacientes con CCLA fueron tratadas al menos una parte o todas ellas con quimio-radioterapia concurrente.

6.1.1.1.4 Tipos de medidas de desenlace.

Los eventos de interés para la RS fueron aquellos desenlace dependientes del tiempo, es decir, aquellos eventos evaluados a 1, 2, 3, 4 o 5 años o durante el tiempo de seguimiento.

6.1.1.1.4.1 Desenlaces principales.

1. Supervivencia global.
2. Supervivencia causa-específica (mortalidad debida al cáncer).
3. Supervivencia libre de enfermedad.
4. Supervivencia libre de recurrencia a distancia.
5. Supervivencia libre de recurrencia local.
6. Supervivencia libre de recurrencia paraaórtica.

6.1.1.2 Métodos de búsqueda para la identificación de estudios.

6.1.1.2.1 Búsquedas electrónicas.

Las bases de datos consultadas para la realización de la RS fueron:

1. Medline (pubmed), (fecha de la última búsqueda Octubre del 2016).
2. ISI Web Of Knowledge, (fecha de la última búsqueda Octubre del 2016).
3. EMBASE (Ovid), (fecha de la última búsqueda Octubre del 2016).
4. IME (Índice Médico Español), (fecha de la última búsqueda Junio del 2015).
5. Epistemonikos, (fecha de la última búsqueda Junio del 2015).
6. IBECs (isciii), (fecha de la última búsqueda Junio del 2015).
7. LILACS (Literatura Latinoamericana de Información en Ciencias de la Salud), (fecha de la última búsqueda Junio del 2015).

6.1.1.2.1.1 Vocabulario y palabras de texto de la estrategia de búsqueda.

La búsqueda completa utilizando palabras de texto comunes y términos estandarizados asignados por indexadores (parte de un vocabulario controlado o tesoro) se desarrolló basado en las palabras que conformaron la estrategia de búsqueda electrónica. Se combinaron dos criterios: 1) Asesoramiento de las palabras y términos temáticos por parte de expertos. 2) Evaluación de las palabras de texto común y palabras de búsqueda utilizadas en artículos conceptuales (25,30,32,46,49,83–86) y metodológicos sobre modelos de pronóstico (44,87,88), guías para elaborar documentos científicos sobre modelos de pronóstico (47,89,90), artículo de desarrollo de modelos de pronóstico, de validación externa, de puesta a punto y estudio de impacto sobre modelos de pronóstico (50,91–98), revisiones sistemática de modelos de pronóstico (99–105), artículos generales y revisiones sobre el cáncer de cérvix (1,4,8,11,22,106–108), y de artículos sobre búsquedas óptimas para identificar estudios sobre pronóstico o predicción (99,109–113).

La estrategia de búsqueda específica para cada una de las bases consultadas aparece detallada en la página 141.

6.1.1.3 Extracción de datos y análisis.

6.1.1.3.1 Selección de estudios.

La identificación de los estudios se realizó ejecutando en primer lugar la estrategia de búsqueda en cada una de las bases de datos (Apartado 11.1). Los artículos son seleccionados según el título por un solo autor. Las citas seleccionadas son evaluadas por dos revisores de manera independiente e íntegra (título, resumen y artículo). La selección realizada por cada una de los revisores es confrontada, y, posteriormente, evaluada conjuntamente. En caso de existir alguna discrepancia entre los dos revisores, fueron resueltas por un tercer investigador.

La bibliografía de cada uno de los artículos seleccionados fue revisada, así como, la realización de una búsqueda de referencias citadas (21) vía ISI Web Of Knowledge de cada uno de ellos. Esta búsqueda fue aplicada a los artículos sobre modelo de pronóstico en cáncer de cérvix temprano encontrados.

El texto completo de los artículos seleccionados fue obtenido a través de la Biblioteca Virtual del Hospital Universitario 12 de Octubre, mediante solicitud al servicio de la biblioteca o directamente al autor de correspondencia del artículo.

6.1.1.3.2 Extracción y manejo de los datos.

Un único autor realizó la extracción de los datos de los artículos. Para cada artículo, se extrajo el número de sujetos que conforma la muestra, la localización donde se recogieron los datos, el número de centros implicados, los años de recolección de pacientes, el tipo de estudio, el tratamiento recibido por los pacientes y la clasificación FIGO (5), así como los criterios de inclusión y de exclusión de pacientes, características basales al diagnóstico del cáncer y probabilidades de supervivencia detalladas en el artículo. Para cada modelo de

pronóstico de los artículos, se extrajo el número de sujetos donde se desarrolló, las variables que conformaron el modelo, la fórmula o método de cálculo del riesgo, y las medidas de rendimiento de los modelos, discriminación y calibración (31,43). En los casos en los que existió muestra de validación, se extrajo la misma información que para la muestra de desarrollo.

6.1.1.3.3 Evaluación metodológica de los estudios incluidos.

La evaluación metodológica de los contenidos de los artículos se realizó a través del protocolo TRIPOD (47,89). El protocolo TRIPOD es mostrado en el Apartado 11.4. Listado de 22 ítem esenciales para un buen resumen y presentación de la información de un modelo pronóstico. Cada uno de los ítems fue evaluado con una de las siguientes categorías: ítem no aplicable, no reportado, deficiente, cuestionable o correcto.

6.2 Cohorte de pacientes para la validación externa.

Cohorte histórica de 346 pacientes con cáncer de cérvix no operados tratadas entre julio de 1999 y septiembre de 2014 en el Servicio de Oncología Radioterápica del Hospital Universitario 12 de Octubre, Madrid, España (Figura 4(a)). De las 346 mujeres se excluyeron 49 pacientes por las siguientes razones: 12 mujeres con enfermedad metastásica al diagnóstico, 5 con tumores sincrónicos, 6 embarazadas, 6 con HIV, 6 tratadas inicialmente con quimioterapia neoadyuvante, 10 en las que no se pudo administrar platino concurrente, y 4 con histología de cáncer indiferenciado. Finalmente, 297 mujeres con cáncer de cérvix, de los siguientes tipos histológicos: carcinoma de células escamosas, adenocarcinomas y adenoescamosos, no operadas tratadas con quimio-radioterapia concurrente de intención radical con un esquema de platino i.v. semanal, y con estadios FIGO (5): IB1, IB2, IIA, IIB, IIIB y IVA constituyeron la población de estudio.

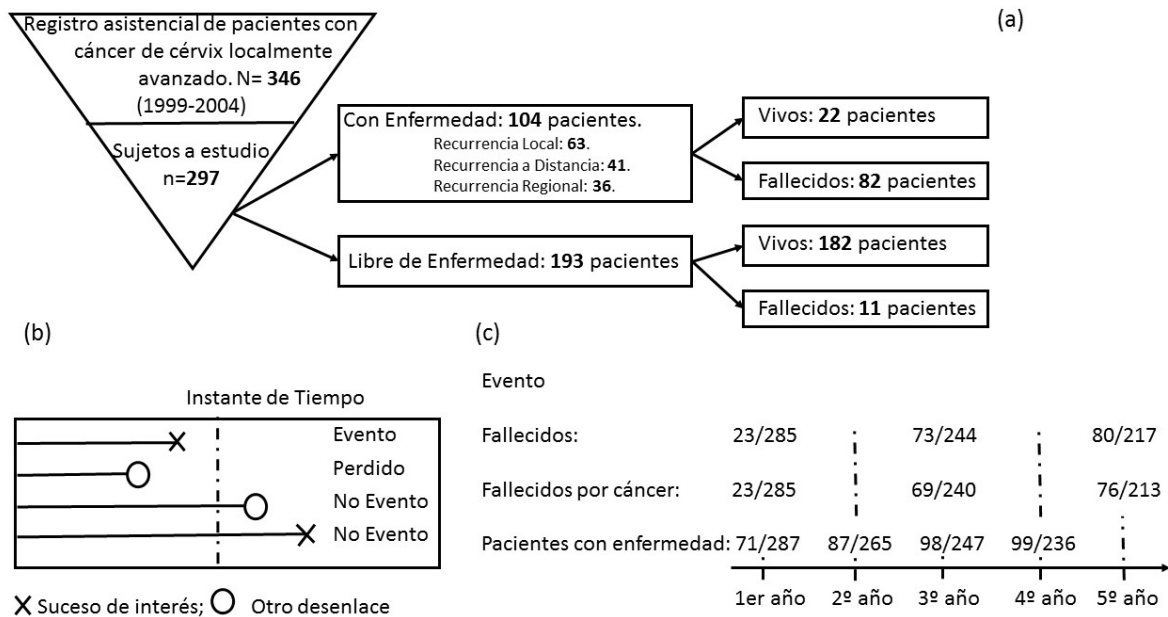


Figura 4. Sujetos y Eventos bajo estudio.

El número de sujetos para validar externamente los modelos pronósticos no fue determinado de manera previa al estudio (26,27,114,115), seleccionando a todas las mujeres registradas en la cohorte histórica que cumplieron los criterios de inclusión (n=297).

6.2.1 Diagnóstico y Tratamiento.

El protocolo de estadificación tumoral utilizado ha sido modificado desde 1999. En primer lugar, el diagnóstico se realizaba a través de una tomografía computarizada (TC), después con TC y/o resonancia magnética (RM) y, posteriormente, con tomografía por emisión de positrones (PET-TC) y/o RM y/o linfadenectomía paraaórtica. A todas las pacientes se les realizó radiografía de tórax y analítica sanguínea completa (sistemático y bioquímica), además se realizó determinación de marcadores tumorales, como el marcador de niveles de antígeno de células escamosas (SCC-Ag) y el marcador Cyfra 21.1. La realización de

rectoscopia o cistoscopia se indicó siempre que existiera sospecha clínica de infiltración o en aquellos tumores \geq IIB.

Las pacientes fueron tratadas con quimio-radioterapia concurrente de intención radical con un esquema de platino i.v. semanal. Desde 1999 hasta fin del estudio.

El tratamiento radioterápico a lo largo de todo el tiempo de estudio, también, ha sufrido modificaciones. La radioterapia externa se ha realizado siempre con cuatro campos conformados e isocéntricos y las dosis administradas han sido entre 45 y 56 Gy. Las simulaciones se realizaron con TC y, a partir del 2009, con PET-TC. Al principio el tratamiento braquiterápico se realizó con Cs 137 de baja tasa de dosis (curietron® de la casa CGR®), siendo los colpostatos utilizados el de Delouche® y, posteriormente, con moldes personalizados de Chassagne. A partir de agosto de 2007, la braquiterapia se realizó con Ir 192 de alta tasa de dosis con moldes de Chassagne sin necesidad de anestesia ni sedación; y además, hemos pasado de prescribir la dosis a puntos ICRU 38 (A, B, rectal y vesical), a prescribir al volumen blanco clínico y a restringir la dosis a los 2 cc de recto, sigma y vejiga).

En los últimos años del estudio, hemos realizado sobreimpresión sobre los ganglios patológicos por imagen (mayores de 1,5 cm de diámetro) durante el tratamiento del campo pélvico o del campo pélvico o paraórtico en forma de boost concomitante hasta 55 Gy (5x220, EQD2 \approx 57 Gy). Las pacientes con tumores bulky y/o \geq IIB y/o con ganglios pélvicos positivos por imagen recibieron radioterapia extendida a paraórticas. En el momento actual, la linfadenectomía paraórtica de estadificación es utilizada en pacientes con PET-TC negativo tanto en pelvis como en paraórticas y sin diseminación a distancia.

6.2.2 Seguimiento de las pacientes.

A los 3 meses de la finalización de la radioterapia, las pacientes tuvieron una consulta, donde se realizó una exploración física, incluyendo, exploración ginecológica y valoración de las pruebas de imagen: TC y/o RM y/o PET-TC y radiografía de tórax. Posteriormente se llevó a cabo un seguimiento de las pacientes cada 3 meses durante los 2 primeros años, cada 6 meses durante los 3 años siguientes y una vez al año a partir del 5 año. En cada visita se realizó un examen físico general, una exploración ginecológica, valoración de las pruebas analíticas y evaluación de la toxicidad tardía de todos los procedimientos empleados. Cada 6 meses, los cinco primeros años y, posteriormente, una vez al año se les realizó una radiografía de tórax y TC de abdomen y pelvis.

El seguimiento de las pacientes para el presente estudio finalizó en Marzo del 2015.

6.2.3 Tratamiento de la información.

Para la realización de esta tesis se utilizó una base de datos anonimizada completamente, por el director de tesis JFPR que es el médico que trató a las pacientes, por lo que el autor de la tesis y el resto de directores desconocían la identidad de las pacientes participantes. No se dispuso de ningún dato de carácter personal, por lo que no se solicitó consentimiento informado a la paciente. Sin embargo, el estudio se presentó en el Comité Ético de Investigación Clínica (CEIC) del Hospital Universitario 12 de Octubre (Nº CEIC: 15/210) para su valoración obteniendo un dictamen favorable a la realización de dicho proyecto en el Centro (Apartado 11.2).

En este estudio, se cumplieron los principios éticos de la declaración de Helsinki y la ley orgánica de protección de datos - LOPD 15/1999.

6.2.3.1 Variables recogidas.

La información de los pacientes disponible en la base de datos fue: nacionalidad, estadio FIGO (5), tamaño del tumor, afectación de los ganglios linfáticos (negativo, afectación en los ganglios pélvicos y/o paraaórticos), tipo histológico (carcinoma de células escamosas, adenocarcinoma y el carcinoma adenoescamoso), escala ECOG (0, 1, 2 y 3), grado histológico del tumor (sin datos, G1, G2, G3), valor máximo estandarizado de captación (SUVmax), metástasis (sí, no), edad al diagnóstico, hemoglobina, antígeno del carcinoma de células escamosas (SCC-Ag), Invasión parametrial (sí, no), invasión de las paredes del recto o de la vejiga (sí, no), hidronefrosis (sí, no), fecha al inicio del tratamiento, fecha a la finalización del tratamiento, fecha de fallecimiento, estado vital (muerto, vivo), causa muerte (cáncer, otro motivo), fecha de recidiva y localización de la recidiva (local, a distancia y paraaórtica).

En general se trabajó con la información disponible, y, por tanto, no se aplicó ninguna técnica de imputación de datos perdidos o ausentes.

6.2.4 Validación externa.

Los modelos de pronóstico validados externamente fueron aquellos identificados por la RS, donde una parte o todas las pacientes de la muestra de desarrollo fueron tratadas con quimio-radioterapia concurrente (10,19).

6.2.4.1 Evento tiempo dependiente.

Definimos evento tiempo dependiente como aquel suceso de interés para el estudio que aparece en las mujeres a riesgo desde el diagnóstico del cáncer hasta un instante concreto del tiempo (116). Por tanto, para evaluar el desenlace se generó una variable dicotómica (Figura

4 (b)) donde se consideró: 1) Evento: Mujeres con el desenlace de interés durante el periodo fijado; 2) No evento: Pacientes con seguimiento superior al instante concreto del tiempo independientemente que posteriormente lo sufrieran; 3) Pérdidas para el análisis: Mujeres con seguimiento inferior al instante de interés sin presentar el evento, debido a la pérdida de seguimiento, por no completar el seguimiento (por ejemplo, mujer diagnosticada en 2014 y evaluar muerte a 5 años), o por evento competitivo con el de interés (por ejemplo, muerte cuando se evalúa metástasis a distancia).

Los eventos tiempo-dependientes de interés para la validación externa fueron:

1. Mortalidad global a 1, 3 y 5 años. Pacientes con seguimiento inferior al instante especificado fueron considerados perdidos.
2. Mortalidad causa debida al cáncer a 3 y 5 años. Pacientes con seguimiento inferior al instante especificado debido a fallecer por otra causa, por pérdida de seguimiento o falta de seguimiento fueron eliminados de los análisis.
3. Recurrencia de la enfermedad a 1, 2, 3 y 4 años.
4. Metástasis a distancia a 5 años.
5. Metástasis paraaórtica a 4 años.

En los puntos 3, 4 y 5, las pacientes con seguimiento inferior al instante especificado debido a fallecer, por perdida de seguimiento o falta de seguimiento fueron eliminados de los análisis.

6.2.4.2 Relación entre las muestras de desarrollo y la muestra de validación.

Las características basales de los sujetos de cada uno de los modelos de pronóstico fueron resumidas y homogeneizadas, expresándose la información según aparece en el artículo: frecuencia absoluta y relativa en caso de ser variables categóricas, media o mediana y desviación estándar o rango intercuartílico o recorrido (mínimo y máximo) en caso de ser

variables cuantitativas. En caso de existir la información, la muestra de validación fue presentada en la misma forma que la muestra de desarrollo.

La distribución de cada una de las características de los sujetos al diagnóstico entre las muestras de desarrollo y la muestra de validación fue comparada (48) mediante el test de Student o de la chi-cuadrado, según naturaleza de las variables, ambos test para muestras independientes.

6.2.4.3 Estimación de las probabilidades predichas por los modelos.

La estimación del riesgo para cada individuo (73) a través de los modelos fue obtenida por diferentes métodos: 1) En caso de existir página web (Tabla 5), a través de la introducción de las características que conforman el modelo en las calculadoras online (41,117); 2) Empleando la herramienta gráfica o nomograma (39,40) propuesta en cada uno de los artículos (118–121), después de haber contactado con los autores de correspondencia vía correo electrónico solicitándoles el modelo explícito, o los valores exactos del nomograma, o la función de supervivencia estimada para los perfiles junto a los coeficientes de regresión (88). Solo un autor de correspondencia proporcionó la información exacta del nomograma (119); 3) Para las estratificaciones del riesgo, se utilizó la probabilidad estimada descrita en el artículo original para cada perfil de sujetos (41,122).

S Polterauer (117)

<http://cemsis.meduniwien.ac.at/en/kb/science-research/software/clinical-software/cervical-cancer/>

S Kang (41)

<http://www.kgog.org/nomogram/kgog1024.html>

Tabla 5. Calculadora online.

En el apartado 1 y 2, los perfiles de los sujetos y la correspondiente probabilidad fueron codificados en un programa en R (123).

6.2.4.3.1 Variables diferentes para la estimación de las probabilidades predichas por los modelos.

Las siguientes variables presentaron diferencias entre la base de datos de validación y la base de datos de desarrollo. A continuación, se presenta la codificación realizada en las variables de la base de validación para poder utilizar cada uno de los modelos de pronóstico (124):

En el modelo de SH Shim (119), la variable afectación de los ganglios linfáticos evaluado por cualquier método (TC, RM, PET-TC o linfadenectomía paraaórtica) en la base de validación fue utilizada para generar la variable estado del ganglio paraaórtico mediante RM necesaria en el modelo de pronóstico.

En las dos estratificaciones del riesgo (libre de enfermedad y paraaórtica) de JA Liang (122), la variable afectación de los ganglios linfáticos recogida en la base de validación fue utilizada para crear la variable estado del ganglio pélvico. En este caso, las categorías “negativa” o “paraaórtico con o sin ganglio pélvico afectado” fueron asignadas a estado del ganglio pélvico igual a “negativo”, y la categoría “pélvico” fue considerada como “positivo”.

En el modelo de S Kang (41), la variable afectación de los ganglios linfáticos evaluado por cualquier método (TC, RM, PET-TC o linfadenectomía paraaórtica) fue utilizada para crear las variables estado del ganglio pélvico y estado del ganglio paraaórtico por PET. La categoría “negativa” conformó “negativo” en las dos variables, la categoría “pélvico” se asignó como “positivo” en la variable estado del ganglio pélvico y “negativo” en la variable estado del ganglio paraaórtico. La categoría “paraaórtico con o sin ganglio pélvico afectado” fue asignada como “negativo” en la variable estado del ganglio pélvico y “positivo” en la variable estado del ganglio paraaórtico.

En el modelo de EA Kidd (120), las tres variables del modelo (tamaño del tumor, ganglio linfático mayor y SUVmax del tumor de cérvix) son recogidas por PET, no siendo de esta manera en la cohorte utilizada para la validación externa. La variable ganglio linfático mayor con cuatro categorías (negativo, pélvico, paraaórtico, supraclavicular) fue generada con la variable afectación de los ganglios linfáticos evaluado por cualquier método (TC, RM, PET-TC o linfadenectomía paraaórtica).

En el modelo de S Polterauer (117), se asignó el valor del ganglio linfático positivo por cada 50 negativos a todas las mujeres debido a la ausencia de esa información en la cohorte de validación.

En los dos modelos de PG. Rose (121), la variable nacionalidad fue utilizada para generar la variable raza. De esta manera, la categoría “Filipinas” de la variable nacionalidad se consideró como raza “Asiática”, las categorías “Bolivia”, ”Brasil”, “Chile”, “Colombia”, “Cuba”, “Ecuador”, “Paraguay”, “Perú”, “República Dominicana” y “Uruguay” como “Hispanos”, las categorías “Alemania”, “Bulgaria”, “Polonia”, "España", "Portugal", “Rumania” y “Ucrania” como raza “Blanca”, las categorías “Angola”, “Camerún”, “Costa Marfil”, “Guinea Bisau” y “Guinea Ecuatorial” como raza “Negra” y la categoría “Marruecos” como raza “Otras”. Para la variable estado del ganglio pélvico se empleó la variable afectación de los ganglios linfáticos donde las categorías “negativa” o “paraaórtico con o sin ganglio pélvico afectado” fueron asignadas a estado del ganglio pélvico “negativo”, y la categoría “pélvico” fue considerada como “positivo”. Finalmente, se consideró que los hallazgos en los ganglios paraaórticos fueron encontrados mediante el empleo de radiología.

6.2.4.4 Cuantificación de la capacidad predictiva de los modelos pronóstico.

Las medidas de evaluación de los modelos de pronóstico estimadas fueron medidas de discriminación, calibración y rendimiento global (31,43,89,115,125). Su cálculo se realizó a través del paquete rms (126) de R (123). Los intervalos de confianza para las medidas de evaluación (127) se estimaron a través de la técnica bootstrap (31), simulando 1000 muestras con 297 sujetos similares a la muestra externa y calculando los intervalos de confianza al 95% bajo la aproximación normal. Se utilizó el paquete boot (128,129) de R (123) fue utilizado.

Discriminación: Se define discriminación como la capacidad del modelo pronóstico para diferenciar entre aquellas mujeres que desarrollaron el evento de aquellas que no lo sufrieron. La discriminación se estimó con el índice C, idéntico al área bajo la curva de rendimiento diagnóstico (81,130) o curva ROC (“Receiver Operating Characteristic”) cuando el desenlace es de naturaleza dicotómica (31). El índice C refleja la probabilidad que para cualquier par de individuos, evento y no evento, el modelo asigne una mayor probabilidad al paciente con el desenlace que al paciente sin él (131).

$$\hat{C} = \frac{1}{n_E n_N} \sum_{j=1}^{n_E} \sum_{i=1}^{n_N} S(x_E, x_N) \quad \text{Ecuación 14}$$

donde:

$$S(x_E, x_N) = \begin{cases} 1, & \text{si } x_E > x_N \\ 1/2, & \text{si } x_E = x_N \\ 0, & \text{si } x_E < x_N \end{cases} \quad \text{Ecuación 15}$$

siendo:

n_E el número de sujetos con evento.

n_N el número de sujetos sin eventos.

X_E el valor de probabilidad para un sujeto con eventos.

X_N el valor de probabilidad para un sujeto sin eventos.

El mínimo valor del índice C es 0,5 e indica que las probabilidades son completamente aleatorias entre eventos y no eventos, por otro lado, el valor máximo del índice C es 1, buena discriminación, por tanto, el modelo ofrece probabilidades más altas a los eventos que a los no eventos.

Calibración: La calibración refleja el acuerdo entre el riesgo predicho por el modelo y el observado en la muestra bajo estudio. La calibración fue estudiada a través de los gráficos de calibración y los coeficientes de calibración (intersección y pendiente de recalibración). La representación gráfica de calibración ofrece las predicciones del modelo en el eje X, y la frecuencia relativa del evento en el eje Y obtenidas a través de técnicas de suavizado (43). Se añade a la figura la línea diagonal que representa el perfecto acuerdo. Los coeficientes de calibración fueron obtenidos a través del modelo logístico de recalibración (43,69):

$$\text{logit}(Y_i) = \alpha + \beta_{\text{cal}} \text{lp}_i \quad \text{Ecuación 16}$$

donde:

Y_i es el estado del paciente i-ésimo.

α es el valor de la intersección cuando $\beta_{\text{cal}}=1$.

β_{cal} es el valor de la pendiente de recalibración.

lp_i es la predicción lineal del pacientes i-ésimo ($\text{lp}_i=X_i\beta$).

Un modelo se dice que está bien calibrado cuando $a=0$ y $b=1$. Los modelos mal calibrados no fueron actualizados (43,49,50).

Medidas de rendimiento global: Se utilizaron dos medidas de rendimiento global, la puntuación de Brier y la R^2 de Nagelkerke.

La puntuación de Brier mide las diferencias entre el riesgo predicho y los eventos observados al nivel de individuo (21), calculando la media al cuadrado de las diferencias entre predicción y evento. El valor de 0 indica acuerdo perfecto.

$$\text{Puntuación de Brier} = \frac{1}{n} \sum_{i=1}^n (P_i - Y_i)^2 \quad \text{Ecuación 17}$$

donde:

n es el número de sujetos.

P_i es la probabilidad predicha por el modelo para el sujeto i -ésimo.

Y_i es la respuesta observada del sujeto i -ésimo.

La estimación R^2 de Nagelkerke evalúa la fuerza predictiva de un modelo a través de la razón del logaritmo de la verosimilitud entre la importancia de la variables que conforman el modelo y el modelo sin ninguna información (31).

$$R_N^2 = \frac{1 - \exp(-LR/n)}{1 - \exp(-L^0/n)} \quad \text{Ecuación 18}$$

donde:

n es el número de sujetos.

LR es el logaritmo del estadístico global de la razón de verosimilitud para el modelo completo.

L^0 es -2 por el logaritmo de la verosimilitud del modelo vacío.

6.2.4.5 Análisis de la curva de decisión.

Los análisis de la curva de decisión (72) se realizaron para aquellos modelos que estaban correctamente calibrados (69), es decir, intervalo de confianza para el valor de la intersección y de la pendiente de recalibración que contuvieran respectivamente a 0 y a 1, y con capacidad discriminativa (43,72), límite inferior del intervalo de confianza para el área ROC superior a 0.5. Se identificó el rango de probabilidades en las cuales los modelos de pronóstico seleccionados fueron de valor según el beneficio neto (Ecuación 13) del análisis de la curva de decisión (72). Estrategia (todos positivos para el desenlace, todos negativos o, positivo o negativo según umbral de probabilidad) con mayor beneficio neto es la mejor.

El análisis se presentó junto con la gráfica de la curva de decisión (Figura 5), que presenta el beneficio neto en el eje vertical y los umbrales de probabilidad sugestiva en el eje horizontal (72,80). Se adjuntó la información de la caracterización (Apartado 12. Publicación) del modelo pronóstico (81) a través de los umbrales de decisión, incluyendo sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo y beneficio neto (Ecuación 13) según estrategia.

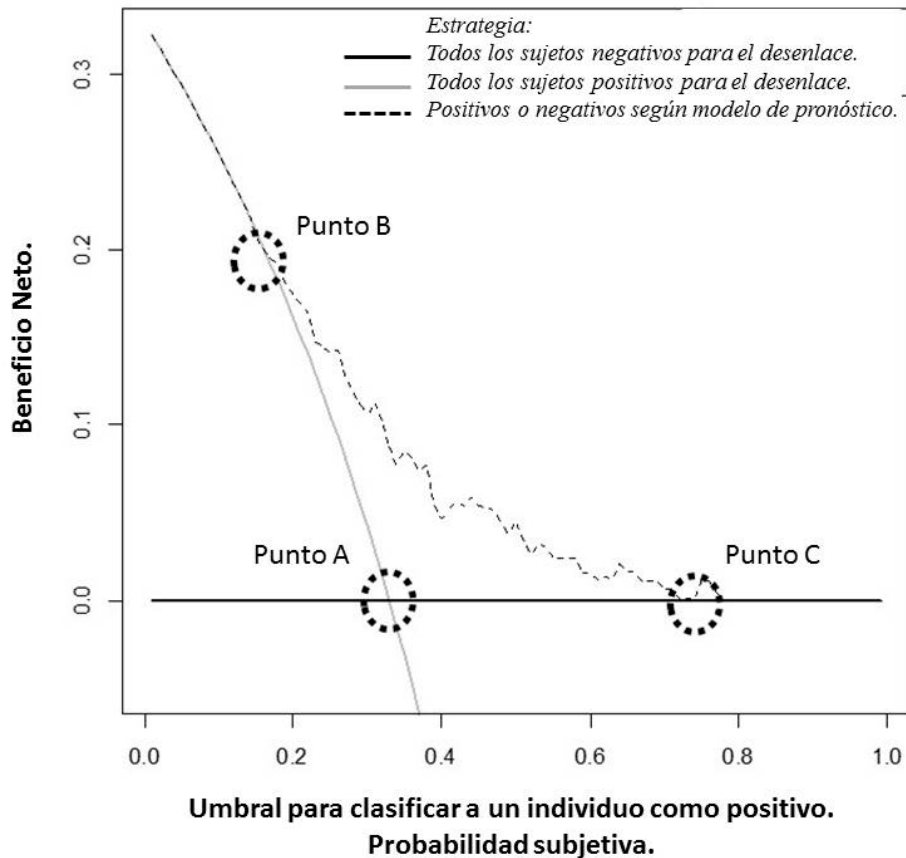


Figura 5. Análisis de la Curva de Decisión.

6.2.4.5.1 Umbrales de relevancia en el análisis de la curva de decisión.

Los umbrales de relevancia en el análisis de la curva de decisión son determinados a través de igualar el beneficio neto entre dos estrategias (72,78). A continuación, se detallan los puntos de interés (Figura 5) siguiendo el desarrollo metodológico de SG Pauker (1980) (78).

Punto A: El beneficio neto de la estrategia de considerar a todos los sujetos como positivos para el desenlace (BN_1) es igual al de la estrategia de catalogar a todos los sujetos como negativos (BN_2). La decisión de clasificar a todos los sujetos como positivos (Tabla 3) es una estrategia permisiva, debido a que se necesita una mínima creencia en la posibilidad del

suceso para indicarlo como positivo, frente a la decisión de clasificar a todos los sujetos como negativos (Tabla 4), estrategia restrictiva.

Ambas estrategias coinciden, cuando la probabilidad subjetiva es igual a la prevalencia del suceso (132) o desenlace de interés (Ecuación 19).

$$BN_1 = \frac{a + c}{N} - \left(\frac{p_u}{1 - p_u} \right) \frac{b + d}{N}$$

$$BN_2 = \frac{0}{N} - \left(\frac{p_u}{1 - p_u} \right) \frac{0}{N} = 0$$

$$BN_1 = BN_2$$

Ecuación 19

$$\frac{a + c}{N} - \left(\frac{p_u}{1 - p_u} \right) \frac{b + d}{N} = 0$$

$$p_u = \frac{a + c}{a + c + b + d}$$

Punto B: El beneficio neto de la estrategia de considerar a todos los sujetos como positivos para el desenlace (BN_1) es igual a la estrategia de clasificar (positivo/negativo) según modelo (BN_3). La Ecuación 20 compara el beneficio neto obtenido de la Tabla 2 (clasificación según modelo) con respecto a la Tabla 3 (clasificación de todos los sujetos como positivo). Ambas estrategias coinciden cuando la probabilidad subjetiva es igual a 1 menos el valor predictivo de un resultado negativo (79,132).

$$BN_1 = \frac{a + c}{N} - \left(\frac{p_u}{1 - p_u} \right) \frac{b + d}{N}$$

Ecuación 20

$$BN_3 = \frac{a}{N} - \left(\frac{p_u}{1 - p_u} \right) \frac{b}{N}$$

$$BN_1 = BN_3$$

$$\frac{a+c}{N} - \left(\frac{p_u}{1-p_u}\right) \frac{b+d}{N} = \frac{a}{N} - \left(\frac{p_u}{1-p_u}\right) \frac{b}{N}$$

$$p_u = \frac{c}{c+d}$$

Punto C: El beneficio neto de la estrategia de considerar a todos los sujetos como negativos para el desenlace (BN_2) es igual a la estrategia de clasificar (positivo/negativo) según modelo (BN_3). Comparación del beneficio neto obtenido de la Tabla 2 con respecto a la Tabla 4. Ambas estrategias coinciden cuando la probabilidad subjetiva es igual al valor predictivo de un resultado positivo (79,132).

$$BN_2 = \frac{0}{N} - \left(\frac{p_u}{1-p_u}\right) \frac{0}{N} = 0$$

$$BN_3 = \frac{a}{N} - \left(\frac{p_u}{1-p_u}\right) \frac{b}{N}$$

$$BN_2 = BN_3$$

Ecuación 21

$$\frac{a}{N} - \left(\frac{p_u}{1-p_u}\right) \frac{b}{N} = 0$$

$$p_u = \frac{a}{a+b}$$

De estos resultados se desprende que, un modelo de pronóstico tendrá más utilidad clínica cuanto más se separen el valor predictivo de un resultado positivo del complementario del valor predictivo de un valor negativo, es decir, valor predictivo de un resultado positivo igual a 1 y valor predictivo de un resultado negativo igual a 0. Por otro lado, un modelo de pronóstico será no informativo cuando el valor predictivo de un resultado positivo sea igual a

la prevalencia del evento y el valor predictivo de un resultado negativo igual a 1 menos la prevalencia del suceso de interés (133).

6.3 Estructura y síntesis del documento de la tesis doctoral.

El manual Cochrane de revisiones sistemáticas de intervenciones (21) fue utilizado para estructurar la RS, en la medida que se pudo adecuar la estructura del documento a la RS de modelos pronósticos. La guía PRISMA (“Preferred Reporting Items for Systematic Reviews and Meta-Analyses”) (134,135) y TRIPOD (47,89) se empleó para ofrecer de manera transparente la información de la RS y la validación externa, respectivamente.

7 RESULTADOS.

7.1 Descripción de los estudios.

7.1.1 Resultados de la búsqueda.

Los estudios incluidos en el proceso de selección para esta revisión son resumidos en la Figura 6. La estrategia de búsqueda aplicada en cada una de las bases de datos (11.1. Estrategias de búsqueda.) identificó 30733 artículos candidatos en Medline (pubmed), 138770 en ISI Web Of Knowledge, 9439 en EMBASE (Ovid), 274 en IME, 7718 en Epistemonikos, 202 en IBECs (isciii) y 34 en LILACS. Un total de 203 artículos fueron incluidos por el título. La inspección posterior del cuerpo del resumen y la lectura del artículo completo en aquellos casos dudosos, permitió identificar 8 artículos (41,117–122,136–142) que satisfacen los criterios de la revisión.

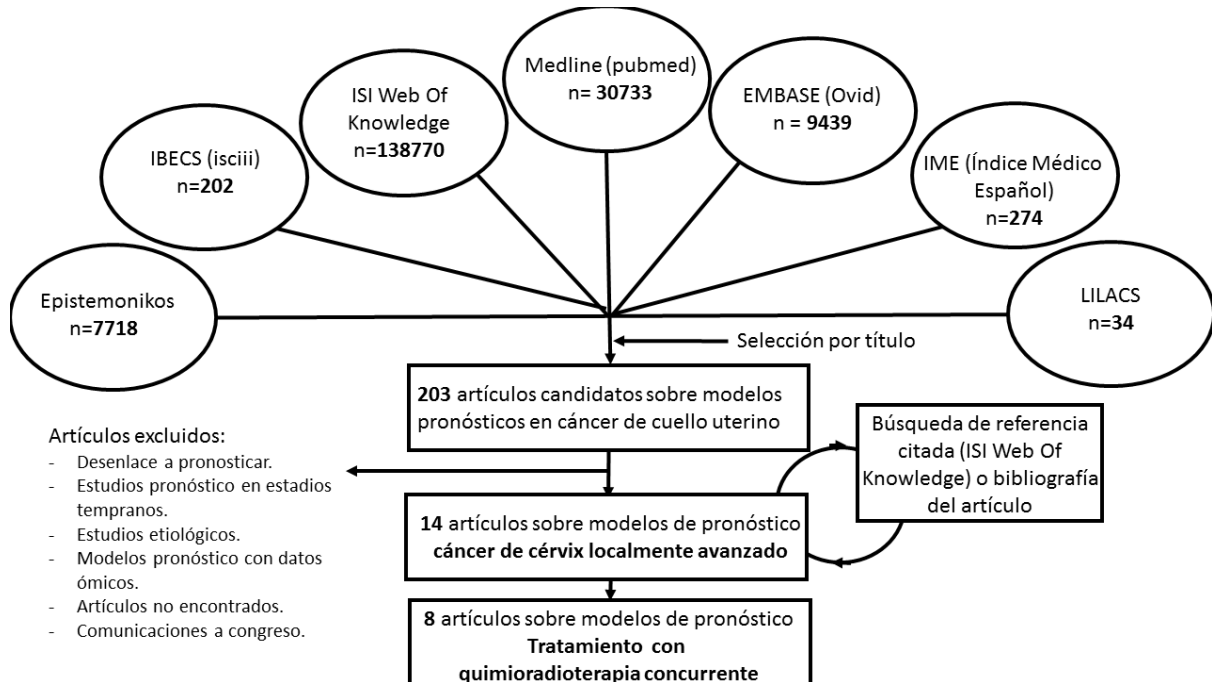


Figura 6. Diagrama de flujo de la revisión sistemática de la literatura.

No se detectó ningún artículo más que cumpliera los criterios de selección a través de la búsqueda de referencia citada vía ISI Web Of Knowledge o búsqueda según la bibliografía recogida en el artículo candidato.

7.1.2 Estudios incluidos.

Los 8 estudios incluidos (41,117–122,142) se detallan en la Tabla 6. Los artículos fueron publicados entre los años 1986 y 2014. Todos los trabajos se basaron en estudios de cohorte, donde la información fue extraída de registros hospitalarios, excepto uno (121), donde la información fue obtenida de seis ensayos clínicos (143–148). En 6 estudios (41,119–122,142), la información se obtuvo de manera retrospectiva, es decir, se identificaron mujeres diagnosticadas con cáncer de cérvix de registros hospitalarios que cumplían los criterios de inclusión y se estudió su evolución. El desarrollo prospectivo de modelos de pronóstico se realizó en 2 estudios (117,118), pacientes reclutados en el diagnóstico y seguidos hacia el futuro.

Autor (Año)		Años	Población; Nº centros D (VE)	Tipo de estudio	Tratamiento principal	Estadio FIGO
X Li (2016)	(142)	2011-2014	China; 1 (-)	Ch-Re	QR	II; III; IVA
SH Shim (2013)	(119)	1998-2008	Corea del Sur; 1 (-)	Ch-Re	QR	IB2; IIA; IIB; IIIA; IIIB
S Kang (2012)	(41)	2001-2009	Corea del Sur; 3 (2)	Ch-Re	QR	IB; IIA; IIB; III; IVA
JY Tseng (2010)	(118)	1999-2006	Taiwan; 1 (-)	Ch-Pr	QR	IIB; IIIA; IIIB; IVA
JA Liang (2008)	(122)	2001-2006	Taiwan; 1 (-)	Ch-Re	QR	IB2; IIA; IIB; III; IVA
S Polterauer (2012)	(117)	1996-2009	Austria; 2 (-)	Ch-Pr	R o Q o QR	IA1; IA2; IB1; IB2; IIA; IIB; IIIA; IIIB; IVA
EA Kidd (2012)	(120)	1998-2008	EEUU; 1 (-)	Ch-Re	R o QR	IB1; IB2; IIA; IIB; IIIA; IIIB; IVA
PG Rose (2015)	(121)	1986-2009	Canada-EEUU; 6 (-)	EC-Re	R o QR	IB; IIA; IIB; IIIA; IIIB; IVA

Tabla 6. Artículos seleccionados en el proceso de revisión sistemática.

Ch: Estudio de Cohorte; EC: Ensayo Clínico; Re: Recogida de la información de manera retrospectiva. Pr: Recogida de la información de manera prospectiva. QR: Quimio-Radioterapia; R: Radioterapia; Qneo: Tratamiento de Quimioterapia neoadyuvante; Q: Quimioterapia; HRcLP: Histerectomía Radical con Linfadenectomía pélvica; (+) Cohorte de desarrollo recogida de manera retrospectiva y la cohorte de validación de manera prospectiva; D: Base de datos de Desarrollo; VE: Base de Validación Externa.

Cinco de los estudios fueron llevados a cabo en hospitales de Asia (Corea del Sur, Taiwan, China y Japón), 2 en América del Norte (EEUU y Canada) y 1 en Europa (Austria). La investigación pronóstica tuvo carácter multicéntrica en 3 de los estudios (41,117,121). El periodo de reclutamiento de los pacientes varió entre 4 y 23 años. La fecha de diagnóstico condiciona el tratamiento de los sujetos en estudio, existiendo 5 estudios (41,118,119,122,142) con quimio-radioterapia concurrente como tratamiento principal y 3 con diferentes tratamientos (117,120,121). En 1 artículo (117) se presentó pacientes con cáncer de cuello uterino en estadios tempranos junto a pacientes con carcinoma cervical localmente avanzado.

La información de los criterios de inclusión o exclusión para cada uno de los estudios es detallada en el Apartado 11.3. De manera general, los criterios de inclusión que aparecen especificados en los artículos son: 1) Pacientes con diagnóstico primario de cáncer de cuello uterino (41,142); 2) Selección de estadios FIGO determinados (41,117,119,121,142); 3) Por tratamiento específico sin que exista operación o tratamiento previo, así como tratamientos posteriores (41,117,119,121,142); 4) Puntuación de ECOG inferiores o iguales a 2 (119,121,142); 5) Pacientes sin enfermedades que afecten a la supervivencia (119,142); 6) Mujeres sin recurrencia del cáncer de cérvix u anomalía antes del tratamiento (142); 7) Según edad (119); y 8) Pacientes con tipo histológico determinado (41,121,142).

En dos estudios no aparecen criterios de inclusión o exclusión (118,122), así como en el artículo (144) que depende de los criterios de selección que aplicó cada uno de los 6 ensayos clínicos.

El número de sujetos candidatos a conformar la muestra del estudio varió entre 163 sujetos y 4221 pacientes. En los 8 artículos (41,117–120,122,142), el número de sujetos se redujo

cuando se aplicaron los criterios de inclusión y exclusión de los propios artículos. Un artículo seleccionó únicamente a un 5% de la muestra general (119). El número de sujetos incluido para cada estudio y aquellos que cumplen con los criterios de inclusión y exclusión aparece reflejado en la Tabla 7, así como la supervivencia para los diferentes desenlaces.

Autor (Año)		Población	Muestra	SG n; p%	SCE n; p%	SLE n; p%	SLEe n; p%
H12O		346	297	70% 5y	71% 5y	64% 5y	
X Li (2016)	(142)	785	300		30; 82% 3y		28 di; 79% 3y
SH Shim (2013)	(119)	4221	209		50; 75% 5y		
S Kang (2012)	(41)	D: 748 VE: 167	D: 434 VE: 115	DE: 74 VE: 21			DE: 92 di; 75% 5y VE: 32 di; 72% 5y
JY Tseng (2010)	(118)	284	Total: 251 IIB: 133 IIIA: 19 IIIB: 75 IIVA: 24		IIB: 62% 5y IIIA: 55% 5y IIIB: 55% 5y IIVA: 16% 5y		
JA Liang (2008)	(122)	163	Total: 148 IB2-IIA: 28 IIB: 83 III-IVA: 37	26	Total: 26 IB2-IIA: 92% 4y IIB: 81% 4y III-IVA: 51% 4y	36*	Total: 18 pa IB2-IIA: 91% 4y IIB: 79% 4y III-IVA: 59% 4y
S Polterauer (2012)	(117)	692	Total: 528 IA: 122 IB: 213 II: 142 III: 37 IV: 14		Total: 76 IA: 99% 5y IB: 89% 5y II: 66% 5y III: 59% 5y IV: 42% 5y		
EA Kidd (2012)	(120)	482	234	74	60	101	

PG Rose (2015)	(144)	2041	Total: 2041		
			IB: 410	IB: 72% 5y	IB: 73% 2y
			IIA: 19	IIA: 72% 5y	IIA: 77% 2y
			IIB: 960	IIB: 63% 5y	IIB: 69% 2y
			IIIA: 23	IIIA: 27% 5y	IIIA: 28% 2y
			IIIB: 566	IIIB: 48% 5y	IIIB: 53% 2y
			IVA: 60	IVA: 35% 5y	IVA: 27% 2y

Tabla 7. Población, sujetos a estudio y supervivencia de los estudios identificados en la revisión sistemática.

SG: Supervivencia Global; SCE: Supervivencia Causa Específica (muerte por cáncer); SLE: Supervivencia Libre de Enfermedad; SLEe: Supervivencia Libre de Enfermedad en localización específica; n: número de sujetos que presentan el desenlace bajo estudio; p%: probabilidad de supervivencia a un tiempo(y) determinado para el evento de interés; D: Base de datos de Desarrollo; VE: Base de Validación Externa; nb: no-bulky; b: bulky; lo: recurrencia local; di: recurrencia a distancia; lo-di: recurrencia local y distancia; pa: recurrencia paraaórtica.

7.1.2.1 Modelos de pronóstico de los estudios incluidos.

Los ocho estudios incluidos conforman un total de 14 modelos de pronóstico, 1 de ellos fue excluido de la revisión debido a que no predicen eventos tiempo dependiente: el modelo de PG Rose (2015) (121) para predecir recurrencia pélvica (si/no). Por tanto, se describen en la revisión 13 modelos de pronóstico (Tabla 8). A partir de 2 modelos de pronóstico (41,120) se muestran estratificaciones del riesgo como una simplificación del modelo original.

En 8 de los 13 modelos de pronóstico con eventos tiempo-dependientes, el evento a considerar fue definido en el texto de forma explícita, y en todos, se explica la manera para calcular el tiempo hasta el evento. Evento y tiempo se indican en 8 artículos. El evento a predecir a través de los modelos de pronóstico fueron: 2 modelos de pronóstico sobre supervivencia global, 5 sobre supervivencia causa específica (muerte debida al cáncer), 3 sobre supervivencia libre de enfermedad y 3 modelos sobre supervivencia libre de enfermedad con localización específica (2 sobre recurrencia a distancia y 1 sobre recurrencia para-aórtica).

La modelización estadística fue la siguiente (no se consideran las 2 estratificaciones del riesgo como simplificación del modelo original): 12 modelos de regresión de Cox y 1 modelo de riesgos competitivos. A partir de los modelos de pronóstico, se construyeron 9 herramientas gráficas, o nomogramas, para estimar la probabilidad de riesgo, 2 de ellas fueron implementadas vía web (los modelos de S Polterauer (117) y S Kang (41)), y 5 estratificaciones del riesgo.

Autor (año)	n	Definición de desenlace	Definición tiempo de seguimiento hasta el desenlace	Tipo de modelo. Herramienta	Área ROC Calibración	
X Li (2016)	(142)	300	Muerte por cáncer	Tiempo desde la completitud del tratamiento hasta la muerte por cáncer.	Modelo de Cox. (estratificación – 3 grupos)	
X Li (2016)	(142)	300	Recurrencia a distancia	Tiempo desde la completitud del tratamiento hasta el primer diagnóstico de recurrencia a distancia.	Modelo de Cox. (estratificación – 3 grupos)	
SH Shim (2013)	(119)	209	Muerte por causa específica	Tiempo desde el inicio de la quimiorradioterapia hasta la muerte	Modelo de Cox. Nomograma	0.69 GC
S Kang (2012)	(41)	D: 434 VE: 115	Recurrencia a distancia definido como recurrencia del tumor en un sitio fuera del campo de radiación pélvica.	Tiempo desde el inicio de la quimiorradioterapia hasta el primer diagnóstico de recurrencia a distancia.	Modelo de riesgos competitivos. web-nomograma (estratificación – 3grupos)	D: 0.70 VE: 0.73 GC
JY Tseng (2010)	(118)	251	Muerte por causa específica	Tiempo desde el inicio de la quimiorradioterapia hasta la muerte causa específica o último seguimiento	Modelo de Cox. Nomograma	0.69 GC
JA Liang (2008)	(122)	148	Libre de enfermedad	Tiempo desde el inicio de la terapia hasta la fecha de última revisión	Modelo de Cox. Estratificación del riesgo (2 grupos)	
JA Liang (2008)	(122)	148	Recurrencia para-áortica	Tiempo desde el inicio de la terapia hasta la fecha de última revisión	Modelo de Cox. Estratificación del riesgo (2 grupos)	
S Polterauer (2012)	(117)	528	Muerte por cáncer	Tiempo desde el diagnóstico hasta la fecha de la muerte por cáncer o último seguimiento	Modelo de Cox. Web-nomograma	0.723 GC

EA Kidd (2012)	(120)	234	No Definido	Supervivencia global	Modelo de Cox. Nomograma	0.658
EA Kidd (2012)	(120)	234	No Definido	Supervivencia Enfermedad Específica	Modelo de Cox. Nomograma	0.739
EA Kidd (2012)	(120)	234	No Definido	Supervivencia libre de recurrencia	Modelo de Cox. Nomograma	0.741
PG Rose (2015)		2041	No Definido	Supervivencia global	Modelo de Cox. Nomograma	0.64 /GC
PG Rose (2015)		2041	No Definido	Supervivencia libre de enfermedad	Modelo de Cox. Nomograma	0.62 /GC

Tabla 8. Modelos de pronóstico para cáncer de cérvix localmente avanzado.

n: número de sujetos para el desarrollo (D) o la validación externa (VE) del modelo de pronóstico; KM: Método de Kaplan-Meier; GC: Gráfico de calibración.

La regla de 10 eventos por cada variable incluida para los modelos de regresión fue cumplida por todos los modelos, excepto por el modelo de Polterauer (117), donde para evitar el sobreajuste se empleó modelos de regresión penalizados.

La capacidad discriminativa y la calibración fue evaluada en 9 (41,117–121) y 6 (41,117–119,121) modelos respectivamente. La capacidad discriminativa varió entre 0.64 y 0.66 para los modelos sobre mortalidad, entre 0.69-0.74 para los modelos de mortalidad por cáncer y entre 0.62-0.74 para los modelos de supervivencia libre de enfermedad. La calibración de los modelos fue representada en los artículos de forma gráfica.

Se realizó validación, interna o externa, de los modelos en 5 de los artículos (41,117,119–121). La validación interna se realizó empleando la técnica bootstrap, en 5 artículos. Uno de los 13 modelos de pronóstico fue validado externamente (41,141).

Los modelos de pronóstico para supervivencia global son dos. Los modelos de pronóstico para supervivencia global utilizan entre 9 y 3 variables para predecir el evento. El tamaño del tumor aparece en ambos modelos (Ver Tabla 9). En general, estadio FIGO más avanzados, tumores más grandes y ganglios linfáticos afectados ofrecen peor pronóstico en los perfiles de los pacientes.

Variables	EA Kidd (2012) (120)	PG Rose (2015) (121)
Estadio FIGO		IB; IIA; IIB; IIIA; IIIB; IVA
Hallazgo ganglio paraaórtico negativo		Patología; Radiología
Ganglio pélvico		Neg; Pos; Desc
Ganglio linfático mayor por PET	Neg; Pélvico; Paraaórtico; Supraclavicular	
Tratamiento		RT+cisplatino; RT +otro
Tamaño tumor	Continua (cm) PET	Continua (cm)
Tipo histológico		Carcinoma de células escamosas; Adenocarcinoma; Adenoescamosas
Raza		Asiática; Hispana; Blanca; Negra; Otras
ECOG		0; 1; 2/3
Grado de diferenciación		Ne; Bueno; Moderado; Pobre
Tumor Cervix SUVmax	Continua	

Tabla 9. Variables de los modelos de pronóstico para supervivencia global.

Los modelos de pronóstico para supervivencia causa específica (muerte debida al cáncer) son 5 (Tabla 10). Los modelos de pronóstico utilizan entre 7 y 3 variables para predecir la supervivencia causa-específica. La variable tamaño del tumor aparece en 4 de los 5 modelos de pronóstico desarrollados, mayor tumor predice un mayor riesgo de fallecer debido al cáncer. Otras variables importantes son el estado de los ganglios (pélvicos, paraaórticos y supraclavicular) y la edad. En el modelo de S Polterauer (117), la edad es considerada protectora frente al evento, más edad menos riesgo de fallecer por cáncer, mientras que en el modelo de JY Tseng (118), más edad indica más riesgo de fallecer por causa-específica.

VARIABLES	X Li (2016) (142)	SH Shim (2013) (119)	JY Tseng (2010) (118)	S Polterauer (2012) (117)	EA Kidd (2012) (120)
Tamaño del tumor		=<4; 4-5; >5	=<4; 4-6; >6	<2; >=2	continua (cm) PET
Estado del ganglio para-aórtico		neg; pos (MRI)			
Metástasis ganglio linfático			Neg; Pélvico; paraaórtico		
Ganglio linfático mayor por PET					Neg; pélvico; paraaórtico; supraclavicular
Razón entre ganglios linfáticos positivos y negativos				continua	
Volumen del ganglio linfático (cm³)	>=3; <3				
Diametro del ganglio linfático (cm)	>=1,5; <1,5				
Adenopatía del ganglio linfático	Neg; pos				
Invasión del parametrio			si; no	si; no	
Estadio FIGO	II; III-IVA			IA; IB; II; III; IV	
Invasión de la vejiga/ recto			si; no		
Tipo histológico		Carcinoma de células escamosas; otros			
Edad			Continua	continua	
SCC-Ag			=<1.5; 1.6- 5; 5.1-15; 15.1>=		
Hidronefrosis			si; no		
Tumor Cervix SUVmax					continua

Tabla 10. Variables de los modelos de pronóstico para supervivencia causa específica.

Tres modelos fueron encontrados para supervivencia libre de enfermedad: 2 modelos de Cox que se representaron como nomogramas, 1 modelo de Cox que permite estratificar el riesgo en dos grupos (Tabla 8). Los modelos para predecir la supervivencia libre de enfermedad utilizan entre 2 y 8 variables. Las variables más importantes en todos los modelos son el estadio FIGO y el tamaño del tumor. Las variables son presentadas en la Tabla 11.

Variables	JA Liang (2008) (122)	EA Kidd (2012) (120)	PG Rose (2015) (121)
Estadio FIGO	IB-IIB; III-IVA		IB; IIA; IIB; IIIA; IIIB; IVA
Estadio del ganglio pélvico	neg; pos		neg; pos
Ganglio linfático mayor por PET		Neg; pélvico; paraaórtico; supraclavicular	
Tamaño del tumor		continua (cm) PET	continua
Tipo de histología			Carcinoma de células escamosas; adenocarcinoma o adenoescamosas
Grado de diferenciación			Ne; bueno; moderado; pobre
Raza			Asiática; hispana; blanca; negra; otras
ECOG			0; 1; 2/3
Tratamiento			RT+cisplatino; RT +otro
Tumor Cervix SUVmax		Continua	

Tabla 11. Variables de los modelos de pronóstico para supervivencia libre de enfermedad.

Fueron 3 los modelos de pronóstico sobre supervivencia libre de enfermedad en localización específica (2 sobre recurrencia a distancia y 1 sobre recurrencia para-). El modelo de S Kang (41) se representa como nomograma asociado con una calculadora vía web (Tabla 5). Este abordaje muestra la estratificación del riesgo en tres grupos. Los otros 2 modelos son estratificaciones del riesgo en 2 y 3. En general, la variable más importante son aquellas

relacionadas con los ganglios linfáticos, mayor riesgo con la presencia de ganglios positivos, y el estadio FIGO, estadios más avanzados asociados a más riesgo (Tabla 12).

VARIABLES	X Li (2016)	S Kang (2012)	JA Liang (2008)
SCC-Ag		continua	
Estado del ganglio pélvico		neg; pos (PET)	neg; pos
Estado del ganglio para-aórtico		neg; pos (PET)	
Volumen del ganglio linfático (cm³)	>=3; <3		
Diametro del ganglio linfático (cm)	>=1,5; <1,5		
Adenopatía del ganglio linfático	Neg; pos		
Tipo histológico		Carcinoma de células escamosas; adenocarcinoma o adenoescamosas	
Estado FIGO			IB-IIB; III-IVA

Tabla 12. Variables de los modelos de pronóstico para supervivencia libre de recurrencia a distancia, paraaórtica y local.

7.1.2.2 Evaluación de la información de los artículos incluidos.

La valoración del conjunto de ítem del TRIPOD (Apartado 11.4) de los estudios seleccionados fue aceptable. No existe ningún artículo que cumpla de forma correcta todos los ítems del TRIPOD. En la Figura 7 se muestran los resultados para los 31 ítems del TRIPOD.

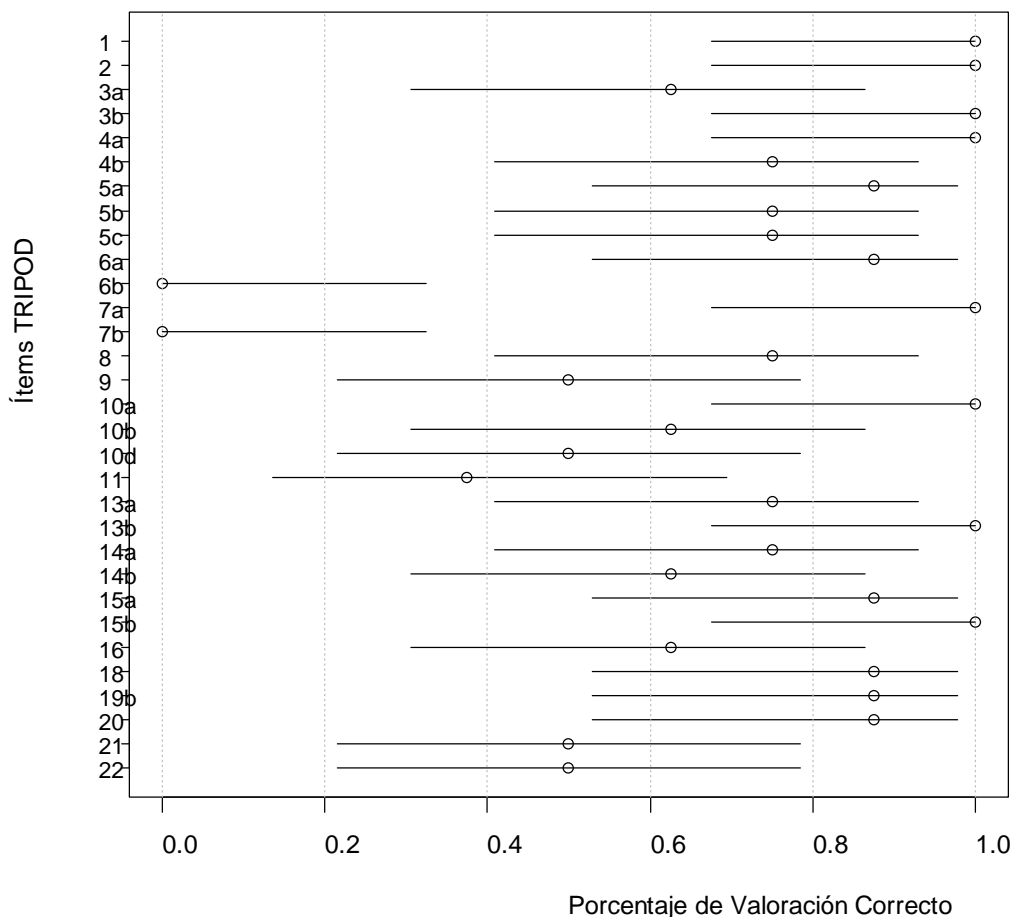


Figura 7. Porcentaje e Intervalo de Confianza del 95% de la valoración correcto para cada ítem del TRIPOD.

En el apartado de "Título y Resumen" (ítems 1 y 2), la valoración correcta en los dos ítem es del 100%.

En el apartado de "Introducción" (ítems 3a y 3b), se encontraron deficiencias en el ítem 3a, siendo la causa mayoritaria la ausencia de referencias de modelos de pronóstico existentes en cáncer de cérvix.

El apartado de "Métodos" (ítems 4a, 4b, 5a, 5b, 5c, 6a, 6b, 7a, 7b, 8, 9, 10a, 10b, 10d y 11) obtuvo las peores puntuaciones en los sub-apartados "Participantes" (ítems 5a, 5b y 5c), "Evento" (6a y 6b), "Tamaño de Muestra" (ítem 8), "Datos Perdidos" (ítem 9) y "Métodos de Análisis Estadísticos" (ítems 10a, 10b y 10d). Destacar del apartado de "Métodos" los

siguientes puntos: Solo en un 75% de los artículos se hizo mención explícita de los criterios de inclusión o exclusión de los participantes (ítem 5b), y en un 87,5% de los artículos se definió claramente el evento a predecir (ítem 6a), incluyendo la definición del tiempo hasta el evento e indicando la fecha inicial y las fechas finales. En 1 artículo, se confunde la palabra modelo para supervivencia global por modelo para supervivencia causa específica. El porcentaje de artículos con la categoría de correcto en el ítem "Tamaño Muestral" (ítem 8) fue del 75%, y el ítem de "Datos Perdidos" (ítem 9) un 50%. El sub-apartado "Métodos de Análisis Estadísticos" obtuvo 100% de valoraciones correcta en el ítem 10a, 62,5% en el ítem 10b y un 50% para el tercer ítem (ítem 10d). En los artículos, el error más frecuente es no especificar las medidas de evaluación de los modelos, así como no indicar los criterios seguidos para determinar el número de grupos de riesgo.

En el apartado de "Resultados" (ítems 13a, 13b, 14a, 14b, 15a, 15b y 16), se encontraron los valores más bajos en los sub-aparatos sobre "Modelo Desarrollado" (ítems 14a y 14b), 75% de valoraciones correctas en el primer ítem y 62,5% en el segundo ítem, y en "Evaluación de los Modelos" (ítem 16), con 62,5% de valoraciones correctas. Las deficiencias son debidas a no incluir medidas de valoración de los modelos, y en caso de aparecer no se ofrece los intervalos de confianza.

En el apartado "Discusión" (ítems 18, 19b y 20), la valoración correcta de los 3 ítems es superior al 80%.

El apartado "Información Adicional", el ítems 21 y el ítems 22 obtuvieron un 50% de valoración correcta.

7.2 Validación externa.

7.2.1 Descripción de la muestra de validación.

El número de sujetos existentes en la cohorte fue de 346. Las mujeres que cumplieron los criterios para ser incluidos en la base de validación externa fueron 297 (Figura 4 (a)). Durante el seguimiento, 93 mujeres fallecieron, probabilidad de supervivencia global a 5 años de 70% (Intervalo de Confianza (IC) 95%: 65-76%), 82 pacientes fallecieron con enfermedad y 22 sin enfermedad, probabilidad de supervivencia causa-específica fue de 71% (IC 95%: 65-76%). En 104 de las 297 mujeres presentaron enfermedad durante el seguimiento, probabilidad de supervivencia libre de enfermedad a 5 años de 64% (IC95%: 58-70%) (Tabla 7). La Figura 8 presenta la localización de la recurrencia de la enfermedad en las pacientes.

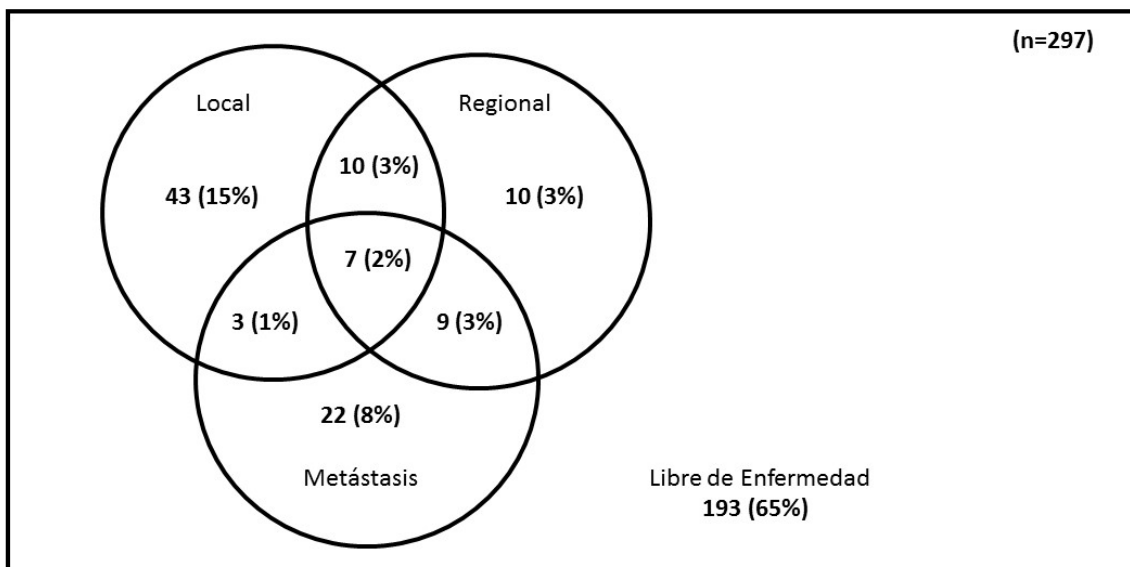


Figura 8. Diagrama de Venn de la recurrencia de las enfermas.

La gráfica de supervivencia global, supervivencia causa específica y libre de enfermedad y estratificada por estadio FIGO para las pacientes de la cohorte de validación externa se

presenta hasta los 5 años en la Figura 9. La información sobre el número de sujetos a riesgo para cada uno de los eventos considerados y por cada año es reflejada al pie de cada figura.

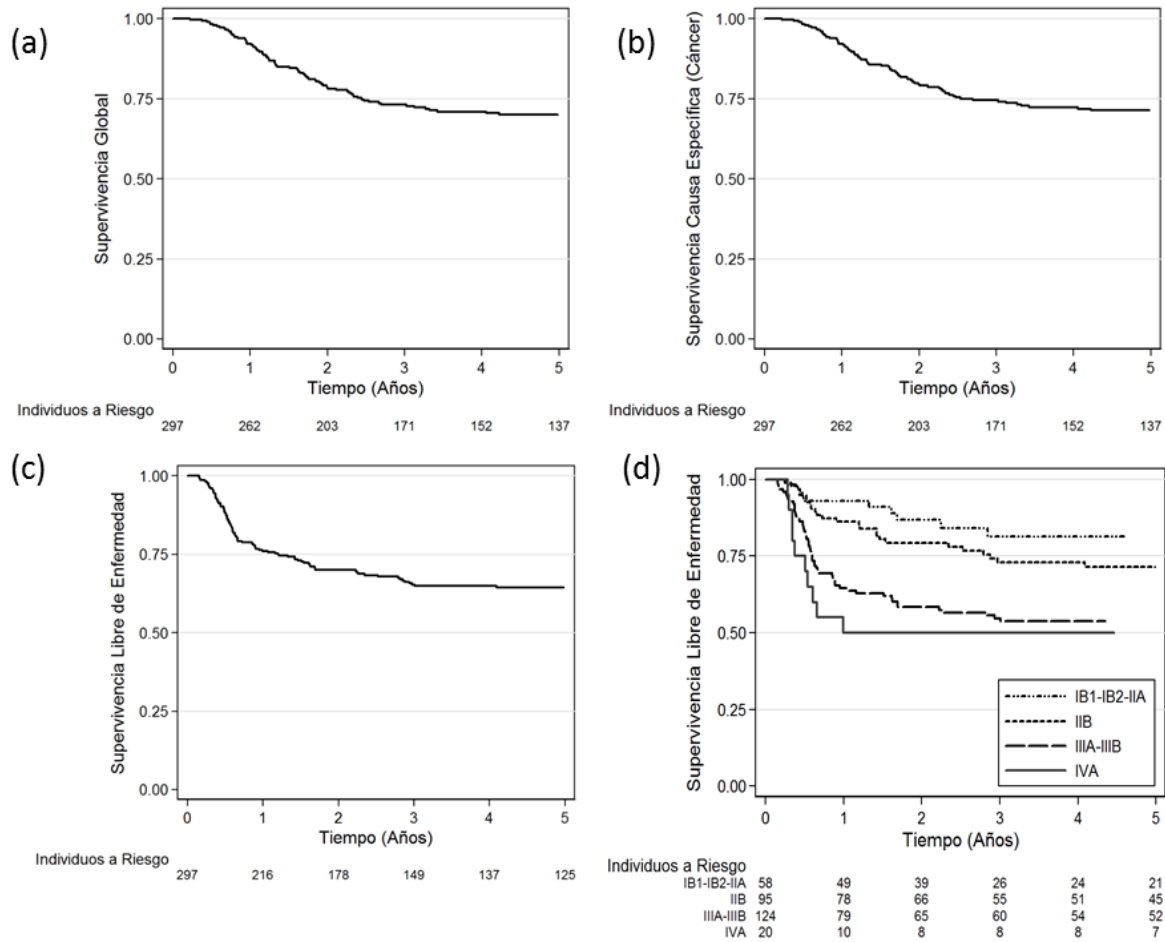


Figura 9. Curva de supervivencia de Kaplan-Meier para la muestra de validación.

7.2.2 Descripción de las muestras de desarrollo.

La información extraída y resumida de los artículos en los que los modelos de pronóstico fueron desarrollados en pacientes donde una parte o todas las mujeres recibieron quimio-radioterapia es recogida en la Tabla 13. Se excluyó de la descripción de la muestra al artículo de X Li (2016) (142) debido a que no fue validado externamente por no disponer de las variables necesarias. La serie de pacientes de la cohorte de validación externa del Hospital

Universitario 12 de Octubre es homogénea respecto a las mujeres de las demás cohortes en cuanto a la edad de las pacientes, con edad promedio de 50 años, el tipo histológico, predominando el tipo de células escamosas por encima de un 79%, y similar ECOG. Se encontraron diferencias significativas (p -valor <0.05) entre el estadio FIGO de las pacientes de la cohorte de validación externa, 49% de pacientes con estadios FIGO de IIIA, IIIB y IVA, frente al resto de cohortes, 40% como máximo. Además existe diferencias significativas, entre la afectación positiva del paramétrio 78% en la cohorte del Hospital Universitario 12 de Octubre frente al 32% cohorte de S Polterauer (2012) (117), así como las cohortes de SH Shim (2013) (119), 87%, y JY Tseng (2010) (118), 87%, estas últimas con dirección semejante. Las cohortes tienen porcentaje similares de mujeres con hidronefrosis, entorno a un 15%, excepto la cohorte de JY Tseng (2010) (118), con 36% de mujeres con hidronefrosis, p -valor < 0.001 . Se encuentran diferencias significativas en la localización de ganglios positivos entre las diferentes cohortes, EA Kidd (2012) (120), SH Shim (2013) (119), S Kang (2012) (41), y la cohorte del Hospital Universitario 12 de Octubre, ganglios pélvicos positivos superiores a un 45% frente a un 28%, respectivamente. Por último, la raza de las mujeres incluidas dependió del lugar donde se realizó el estudio, existiendo 2 estudios con exclusivamente población asiática.

	H12O	SH Shim (2013) (119)	S Kang (2012) D(41)	S Kang (2012) VE (41)	JY Tseng (2010) (118)	JA Liang (2008) (122)	S Polterauer (2012) (117)	EA Kidd (2012) (120)	PG Rose (2015) (121)
N	297	209	434	115	251	148	528	234	2041
Edad	53.9 (13.1)	55 (26-78)	54 (27-78)	54 (32-77)	48.6 (9.3)	54 (31-76)	47.9 (13.5)	52 (24-94)	46.6 [39.0-55.9]
Estadio FIGO:									
IA1-IA2	-	-	-	-	-	-	122 (23.1)	-	-
IB1-IB2-IIA	58 (19.5)	25 (12.0)	72 (16.6)*	8 (7.0)*	19 (7.6)	28 (19.0)*	237 (44.8)	70 (29.9)	433 (21.2)
IIB	95 (32.0)	129 (61.7)	274 (63.1)	71 (62.8)	133 (52.9)	83 (56.1)	118 (22.3)	102 (43.6)	960 (47.0)
IIIA-IIIB	124 (41.8)	45 (21.5)	70 (16.1)	26 (22.6)	75 (29.8)	34 (23.0)	37 (7.0)		589 (28.8)
IVA	20 (6.7)	10 (4.8)	18 (4.2)	10 (8.7)	24 (9.5)	3 (2.0)	7 (1.3)	59 (25.2)	60 (2.9)
IVB	-	-	-	-	-	-	7 (1.3)	3 (1.2)	-
Tipo Histológico									
Cél. Esca	247 (83.2)	190 (90.8)	385 (88.7%)	106 (92.2)	251 (100)	135 (91.2)	421 (79.7)	207 (88)	1811 (88.7)
Adenocar	45 (15.2)	13 (6.2)	29 (6.7%)	4 (3.5)	-	13 (8.8)	86 (16.3)	16 (7)	
Adenoesca.	5 (1.7)	4 (1.9)	17 (3.9%)	2 (1.7)	-	-	21 (4.0)	4 (2)	114 (5.6)
Otras	-	2 (1.0)+	3 (0.7)	3 (2.6)	-	-	-	7 (3)	117 (5.7)
ECOG									
0	24 (8.1)		-	-		-	-	-	1473 (72.1)
1	227 (76.4)		-	-	179 (71.3) ^[<2]	-	-	-	
2	37 (12.5)	209(100) ^[<3]	-	-	72 (28.7)	-	-	-	503 (24.6)
3	9 (3.0)	-	-	-	-	-	-	-	66 (3.2)

SCC-Ag	17.1 (33.6)	6.4 (0-319)		4.9 (0-395)		8.3 (0-402)		168 (66.8) [^{<5}]	-	-	-	-
								83 (33.0) [^{>=5}]				
Tamaño tumor	5.9 (14.6)#	75 (36) [^{<4}]		4.5 (1.6-9)#		5.5 (1.5-9.5)#		49 (19.5) [^{<4}]	11 (7) [^{<4}]	215 (41) [^{<2}]	-	6.0 [5.0-7.0]#
		134(64) [^{>=4}]						202 (80.5) [^{>=4}]	137(93) [^{>=4}]	313(59) [^{>=2}]		
Afectación parametrio		PET										
Negativo	65 (21.9)	28 (13.4)		-		-		33 (13.1)	-	358 (67.8)	-	477 (23.4)
Positivo	232 (78.1)	181 (86.8)		-		-		218 (86.8)	-	170 (32.2)	-	1565 (76.6)
Hidronefrosis												
Negativo	245 (82.5)	174 (83.3)		-		-		160 (63.7)	-	-	-	1490
Positivo	52 (17.5)	35 (16.7)		-		-		91 (36.3)	-	-	-	(89.2)
												181 (10.9)
Ganglio linfático		(1)	(2)	(1)	(2)	(1)	(2)					
Negativo	166 (55.9)	111	180	174	346	56	90	142 (56.6)	-	389 (73.7)	109 (47)	1285
Pélvico	83 (28.0)	98	-	260	-	59	-	71 (28.3)	-	139 (26.3)	125 (53)	(62.9)
Paraaórtico	48 (16.2)	-	29	-	88	-	25	38 (15.1)	-	-	41 (18)	286 (14.0)
Supraclavi.	-	-	-	-	-	-	-	-	-	-	10 (4)	-
												-
Raza												
Blancos	236 (79.7)	-		-		-		-	-	-	-	1242(60.8)
Negros	7(2.4)	-		-		-		-	-	-	-	459 (22.5)
Hispanos	45 (15.2)	-		-		-		-	-	-	-	210 (10.3)
Asiáticos	2 (0.7)	-		434 (100)		115 (100)		-	-	-	-	75 (3.7)
Otros	6 (2.0)	-		-		-		-	-	-	-	56 (2.7)

Tabla 13. Características de las mujeres de los estudios incluidos para la validación externa.

*H12O: Servicio de Oncología Radioterápica del Hospital Universitario 12 de Octubre; D: Base de datos de Desarrollo; VE: Base de Validación Externa; PET: Tomografía por Emisión de Positrones; *: Inclusión de la categoría bulky; +: Células pequeñas. #: Variable expresada en medidas continuas: media (desviación estándar), media (mínimo-máximo) o mediana [rango intercuartílico].*

7.2.3 Validación externa de los modelos de pronóstico.

Los modelos de pronóstico validados externamente, que cumplieron el criterio de ser desarrollados en pacientes donde una parte o todas las mujeres recibieron quimio-radioterapia, fueron 11 de los 20 identificados e incluidos en la RS. Cada uno de los modelos fue evaluado en los tiempos especificados. Se añadió la evaluación de la estratificación del modelo de S Kang (2012)(41). La cuantificación de la capacidad predictiva de los 11 modelos pronósticos para cada uno de los momentos, a través, de la discriminación, calibración y las medidas de rendimiento global, son reflejadas en la Tabla 14. Todos los conjuntos presentaron un número inferior a 100 eventos en la muestra de validación externa. Además los modelos de EA Kidd (2012) (120), con muestra inferior a 100 sujetos, junto con el modelo de S Polterauer (117) para supervivencia causa específica a 1 año y el modelo de S Kang (2012) (41) tuvieron menos de 50 eventos.

	Modelo	Eventos /Muestra	Tiempo	Área ROC	Intersección	Pendiente de recalibración	R²_N	Puntuación de Brier
G	PG Rose (2015) (121)	80/216	5y	0.66	-0.72	0.82	10.40%	0.21
				0.66	-0.72	0.83	10.67%	0.21
				(0.59;0.74)	(-1.04;-0.43)	(0.43;1.25)	(3.01;20.71)	(0.19;0.23)
G	EA Kidd (2012) (120)	9/92	1y	0.67	-0.17	1.01	5.24%	0.09
				0.67	-0.17	1.01	5.74%	0.08
				(0.48;0.86)	(-2.95;2.32)	(-0.30;2.32)	(0.0;16.82)	(0.04;0.14)
G	EA Kidd (2012) (120)	21/53	3y	0.69	0.23	0.88	5.32%	0.23
				0.69	0.25	0.88	5.79%	0.22
				(0.54;0.84)	(-0.81;1.38)	(-0.30;2.17)	(0.0;30.54)	(0.20;0.28)
E	SH Shim (2013) (119)	69/240	3y	0.52	-0.89	0.01	0.01%	0.20
				0.52	-0.91	0.01	0.02%	0.20 (0.18;0.23)
				(0.45;0.60)	(-1.44;-0.41)	(-0.36;0.37)	(0.0;0.30)	
E	SH Shim (2013) (119)	76/213	5y	0.55	-0.49	0.13	0.39%	0.23
				0.55	-0.49	0.14	0.50%	0.23
				(0.47;0.63)	(-0.86;-0.12)	(-0.18;0.53)	(0.0;5.11)	(0.21;0.24)
E	JY Tseng (2010) (118)	75/211	5y	0.64	-1.30	0.98	8.59%	0.21
				0.65	-1.30	1.00	8.42%	0.21
				(0.57;0.72)	(-1.94;-0.83)	(0.51;1.66)	(1.78;18.42)	(0.19;0.23)
E	EA Kidd (2012) (120)	9/92	1y	0.60	-1.37	0.38	1.44%	0.09
				0.60	-1.37	0.40	2.10%	0.09
				(0.38;0.82)	(-3.66;0.77)	(-0.55;1.42)	(0.0; 17.87%)	(0.04;0.14)
E	EA Kidd (2012) (120)	20/52	3y	0.58	-0.04	0.31	1.83%	0.23
				0.57	-0.06	0.33	2.36%	0.23
				(0.41;0.75)	(-1.21;1.14)	(-0.43;1.07)	(0.0;24.87)	(0.17;0.25)

E	S.Polterauer (2012) (117)	23/284	1y	0.65	0.70	1.34	5.02%	0.07
				0.65	0.72	1.34	5.28%	0.07
				(0.55;0.74)	(-1.44;2.99)	(0.42;2.28)	(0.75;12.47)	(0.05;0.10)
E	S.Polterauer (2012) (117)	69/239	3y	0.63	-0.01	0.98	5.52%	0.20
				0.63	0.01	1.00	5.38%	0.20
				(0.55;0.71)	(-0.65;0.64)	(0.37;1.75)	(0.69;13.81)	(0.17;0.22)
E	S.Polterauer (2012) (117)	76/212	5y	0.63	-0.35	0.83	5.37%	0.22
				0.63	-0.36	0.85	5.45%	0.22
				(0.55;0.71)	(-0.67;-0.06)	(0.27;1.48)	(0.58;14.19)	(0.20;0.24)
LE	PG Rose (2015) (121)	87/264	2y	0.70	-0.25	1.19	14.42%	0.20
				0.70	-0.26	1.20	14.53%	0.19
				(0.63;0.76)	(-0.59;0.09)	(0.75;1.71)	(6.02;25.17)	(0.17;0.22)
LE	EA Kidd (2012) (120)	26/93	1y	0.64	-0.24	0.57	4.17%	0.20
				0.64	-0.26	0.57	4.15%	0.19
				(0.51;0.76)	(-1.24;0.82)	(-0.12;1.33)	(0.0;18.27)	(0.15;0.23)
LE	EA Kidd (2012) (120)	32/58	3y	0.60	0.32	0.26	1.23%	0.24
				0.60	0.35	0.30	1.80%	0.24
				(0.45;0.73)	(-0.27;1.34)	(-0.41;1.45)	(0.0;17.73)	(0.20;0.25)
LE	JA Liang (2008) (122)	99/236	4y	0.64	-	-	-	-
				(0.59;0.70)				
Rd	S Kang (2012) (41)	38/163	5y	0.67	0.25	1.00	7.77%	0.17
				0.68	0.28	1.01	7.77%	0.17
				(0.58;0.77)	(-0.67;1.39)	(0.34;1.73)	(1.03;19.64)	(0.13;0.20)
Rd	S Kang (2012) (41)	38/163	5y	0.63	-	-	-	-
				(0.54;0.72)				
Rp	JA Liang (2008) (122)	18/155	4y	0.58	-	-	-	-
				(0.46;0.69)				

Tabla 14. Validación externa de los modelos de pronóstico.

Las celdas contienen el valor estimado, la estimación via bootstrap y su intervalo de confianza al 95% usando la aproximación normal. ROC: Receiver Operating Characteristic; G: Supervivencia Global; E: Supervivencia Causa Específica (causa cáncer); L:Supervivencia Libre de Enfermedad; Rd: Supervivencia libre de Recurrencia a distancia; Rp: Supervivencia libre de Recurrencia paraaórtica; y: instante (año) de Tiempo de evaluación.

Los modelos validados externamente para supervivencia global fueron 2, con 3 evaluaciones del evento-tiempo. El modelo de pronóstico para supervivencia global a 3 años de EA Kidd (2012) (120) tuvo una capacidad discriminativa de 0.69 (IC95%: 0.54-0.84), frente, al modelo de EA Kidd (2012) (120) para 1 año que no mostró capacidad de diferenciación entre los sujetos que fallecieron en 1 año y el resto de individuos. Los dos modelos de EA Kidd (2012) (120) mostraron correcta calibración en el valor de la intersección y la pendiente. El modelo pronóstico para supervivencia global a 5 años de PG Rose (2015) (121) tuvo una capacidad discriminativa de 0.66 (IC95%: 0.59 - 0.74). La capacidad discriminativa se refleja en el gráfico de cajas (Figura 10 (c)). El modelo de PG Rose (2015) (121) 5 años mostró fallos en la calibración de la intersección, valor inferior a 0, es decir, sobreestima, lo que indica probabilidades predichas sistemáticamente mayores a las que corresponden. En la Figura 10 (d), se observa como la curva de calibración siempre es inferior a la línea ideal. En cambio, presentó buenos valores de rendimiento pronóstico, R² de 10.4% y Brier's score de 0.21.

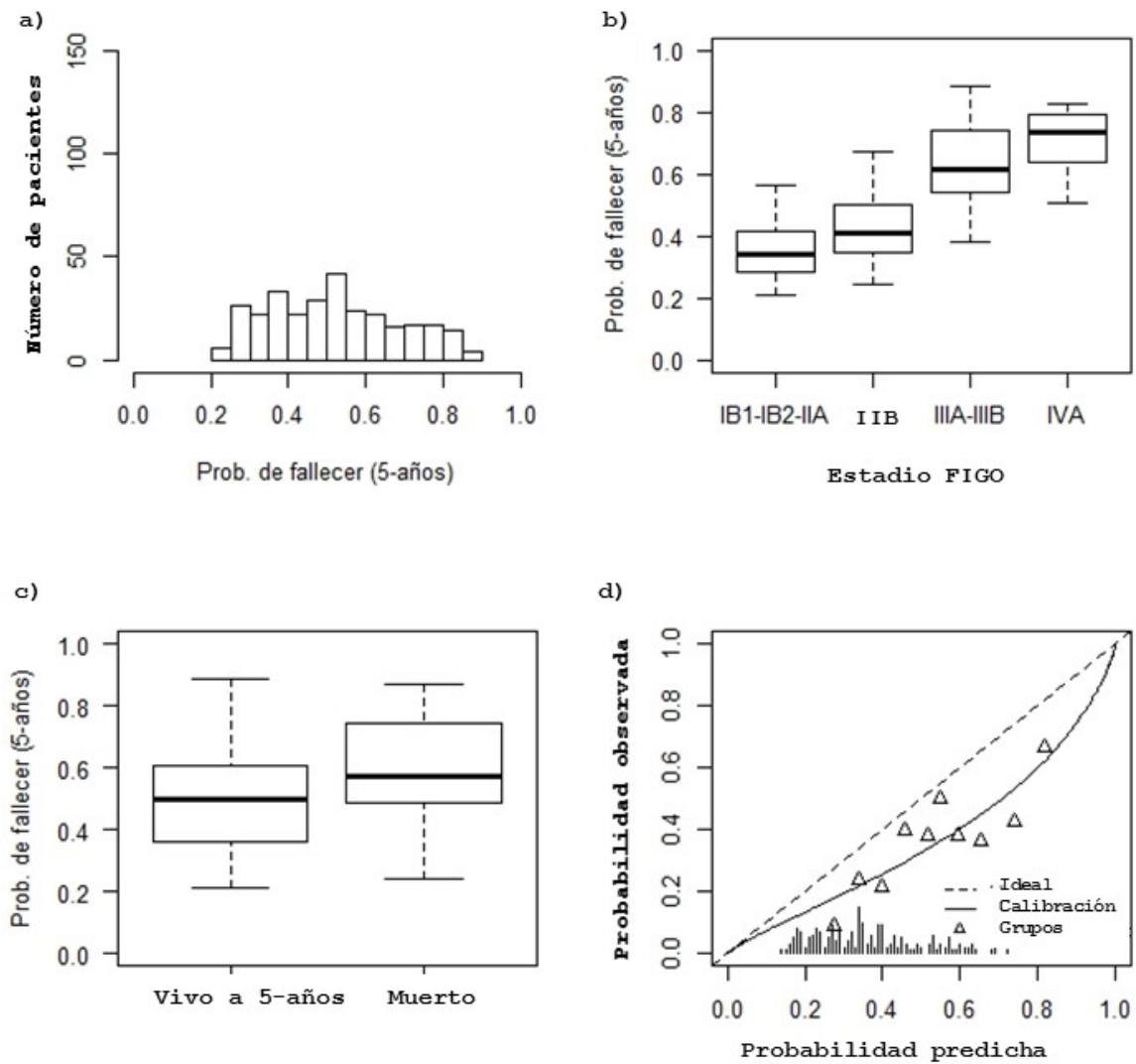


Figura 10. Validación externa del modelo de supervivencia global a 5 años de PG Rose (2015) (121). (a) Distribución de las probabilidades predichas por el modelo. (b) Diagrama de cajas entre la probabilidad predicha y los estadios FIGO agrupados en IB1-IB2-IIA, IIB, IIIA-III B y IVA. (c) Diagrama de cajas entre las probabilidades predichas por el modelo y el evento a predecir, gráfico de discriminación. (d) Gráfico de calibración.

Los modelos de pronóstico para supervivencia causa específica fueron 4, con 8 momentos evento-tiempo (2 evaluados a 1 año, 3 a 3 años y 3 a 5 años). Los 2 modelos de SH Shim (2013) (119) y los 2 modelos de EA Kidd (2012) (120) no discriminan entre las mujeres que sufren la muerte debida al cáncer y el resto. Mientras que los modelos de EA Kidd (2012) (120) están bien calibrados, los 2 modelos de SH Shim (2013) (119) ofrecieron probabilidades predichas más alta de lo observado, sobreestimación. Por otro lado, el modelo de JY Tseng (2010) (118) obtuvo una capacidad discriminativa de 0.64, pero presentó mala calibración en la muestra de validación externa. Los 3 modelos de S Polterauer (2012) (117) a 1, 3 y 5 años presentaron un área bajo la curva superior a 0.63, acompañado de una correcta calibración en el valor de la intersección y la pendiente. Estos 3 modelos fueron adecuados en cuanto a discriminación y calibración. Además explicaban en torno al 5% de la variabilidad del evento de interés. La gráfica de caja para mostrar la discriminación y el gráfico de calibración son mostrados en la Figura 11 (a y b), respectivamente, para el modelo de supervivencia causa específica a 3 años de S Polterauer (2012) (117).

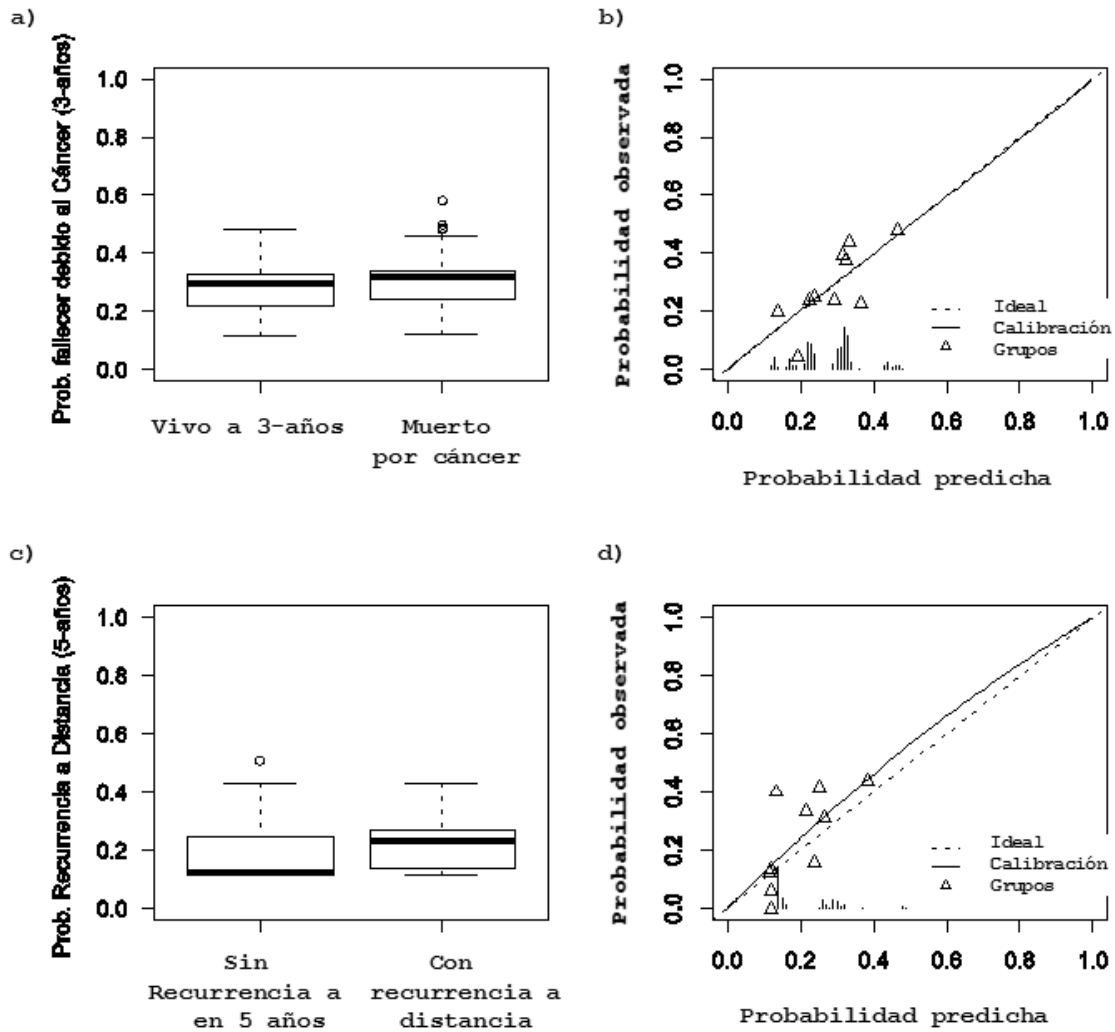


Figura 11. Validación externa del modelo de supervivencia causa específica a 3 años de S Polterauer (2012) (117) [arriba] y del modelo para recurrencia a distancia de S Kang (2012) (41) [abajo]. (a) Diagrama de cajas de las probabilidades predichas por el modelo de S Polterauer (2012) (117) y el evento a predecir, gráfico de discriminación. (b) Gráfico de calibración para el modelo de S Polterauer (2012) (117). (c) Diagrama de cajas de las probabilidades predichas por el modelo de S Kang (2012) (41) y el evento a predecir, gráfico de discriminación. (d) Gráfico de calibración para el modelo de S Kang (2012) (41).

Únicamente, los modelos para supervivencia libre de enfermedad en los momentos de 1 año de EA Kidd (2012) (120) y 2 años de PG Rose (121) tuvieron un área bajo la curva superior a 0.64, destacando el modelo de PG Rose (121) a 2 años con una capacidad discriminativa de 0.70 (IC95%: 0.63-0.76). El modelo de EA Kidd (2012) (120) a 3 años no consiguió diferenciar correctamente a los eventos de los no eventos. Todos los modelos fueron correctamente calibrados, valor de la intersección igual a 0 y pendiente de 1. Destaca las medidas de rendimiento global del modelo de PG Rose (2012) (121) con respecto a los demás modelos, con un R2 de 14.4% y un valor de Brier's score de 0.20. La información gráfica del modelo de PG Rose (2012) (121) se presenta en la Figura 12.

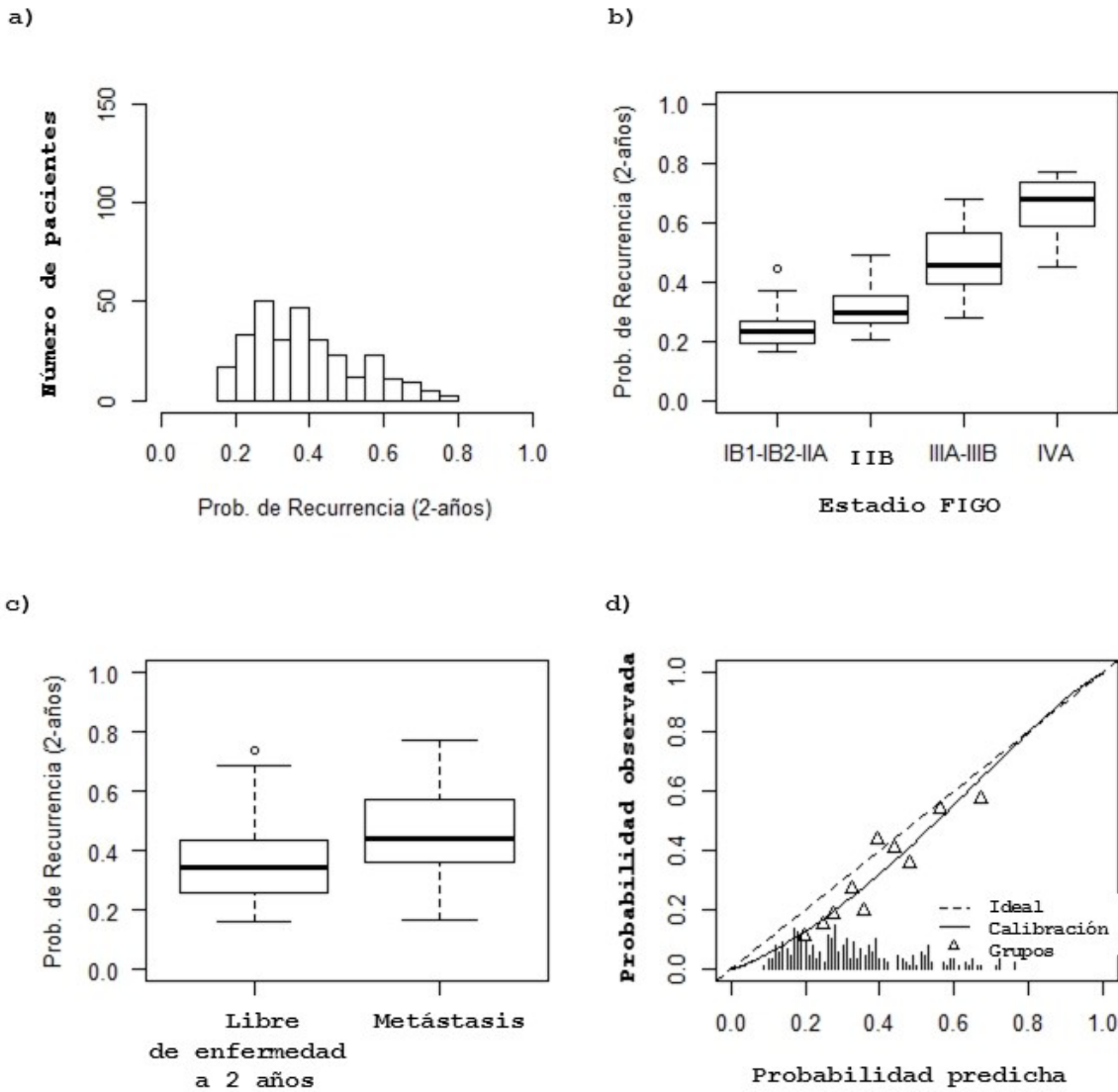


Figura 12. Validación externa del modelo de supervivencia libre de enfermedad a 2 años de PG Rose (2012) (121). (a) Distribución de las probabilidades predichas por el modelo. (b) Diagrama de cajas de la probabilidad predicha por estadios FIGO agrupados en IB1-IB2-IIA, IIB, IIIA-IIIIB y IVA. (c) Diagrama de cajas entre las probabilidades predichas por el modelo y el evento a predecir, gráfico de discriminación. (d) Gráfico de calibración.

El único modelo de pronóstico que existe para recurrencia a distancia es el modelo de pronóstico de S Kang (2012) (41), que ofrece predicciones para 5 años. Este modelo gozó de buena capacidad discriminativa (Figura 11 (c)), 0.67 (IC95%: 0.58-0.77), así como buena calibración (Figura 11 (d)) y adecuadas medidas de rendimiento, R2 de 8% y 0.17 de puntuación de Brier.

Por último, las 2 estratificaciones del riesgo para supervivencia libre de enfermedad a 4 años de JA Liang (2008) (122) y para recurrencia a distancia a 5 años de S Kang (2012) (41) presentaron una capacidad discriminativa de 0.64 y 0.63, respectivamente. La estratificación para predecir recurrencia paraaórtica a 4 años de JA Liang (2008) (122) no discriminó entre los eventos de los no eventos, 0.58 (IC95%: 0.46-0.69).

7.2.4 Análisis de las curvas de decisión.

Las 6 evaluaciones de las predicciones en el tiempo de los 5 modelos que aparecen en la Figura 13, presentaron pequeñas regiones de utilidad clínica debidas al impacto de la baja capacidad discriminativa (área bajo la curva ROC inferiores a 0.70), calibración no perfecta (intersección diferente de 0 y pendiente de recalibración diferente de 1) y el número reducido de eventos (menos de 100 sujetos con el desenlace de interés). Una capacidad discriminativa baja implica que las regiones de probabilidades donde el modelo es de utilidad son estrechas. La figura nos muestra que para el modelo de supervivencia global a 3 años desarrollado por EA Kidd (2012) (120) los umbrales fueron entre 24% y 50% (Figura 13 (a)), para los modelos de supervivencia causa específica de S Polterauer (2012) (117) fueron entre 4% y 11% para 1 años (Figura 13 (c)), de entre 12% y 36% para 3 años (Figura 13 (d)), para el modelo de supervivencia libre de enfermedad a 1 año de EA Kidd (2012) (120) fueron entre 7% y 32% (Figura 13 (e)), para el modelo de supervivencia libre de enfermedad a 2 años de

PG Rose (2012) (121) fueron entre 19% y 58% (Figura 13 (f)), y para el modelo de recurrencia a distancia de S Kang (2012) (41) a 5 años fue entre 12% y 36% (Figura 13 (b)). Las curvas del análisis del beneficio neto se mostraron muy poligonales debido a estimaciones imprecisas debido al bajo número de sujetos.

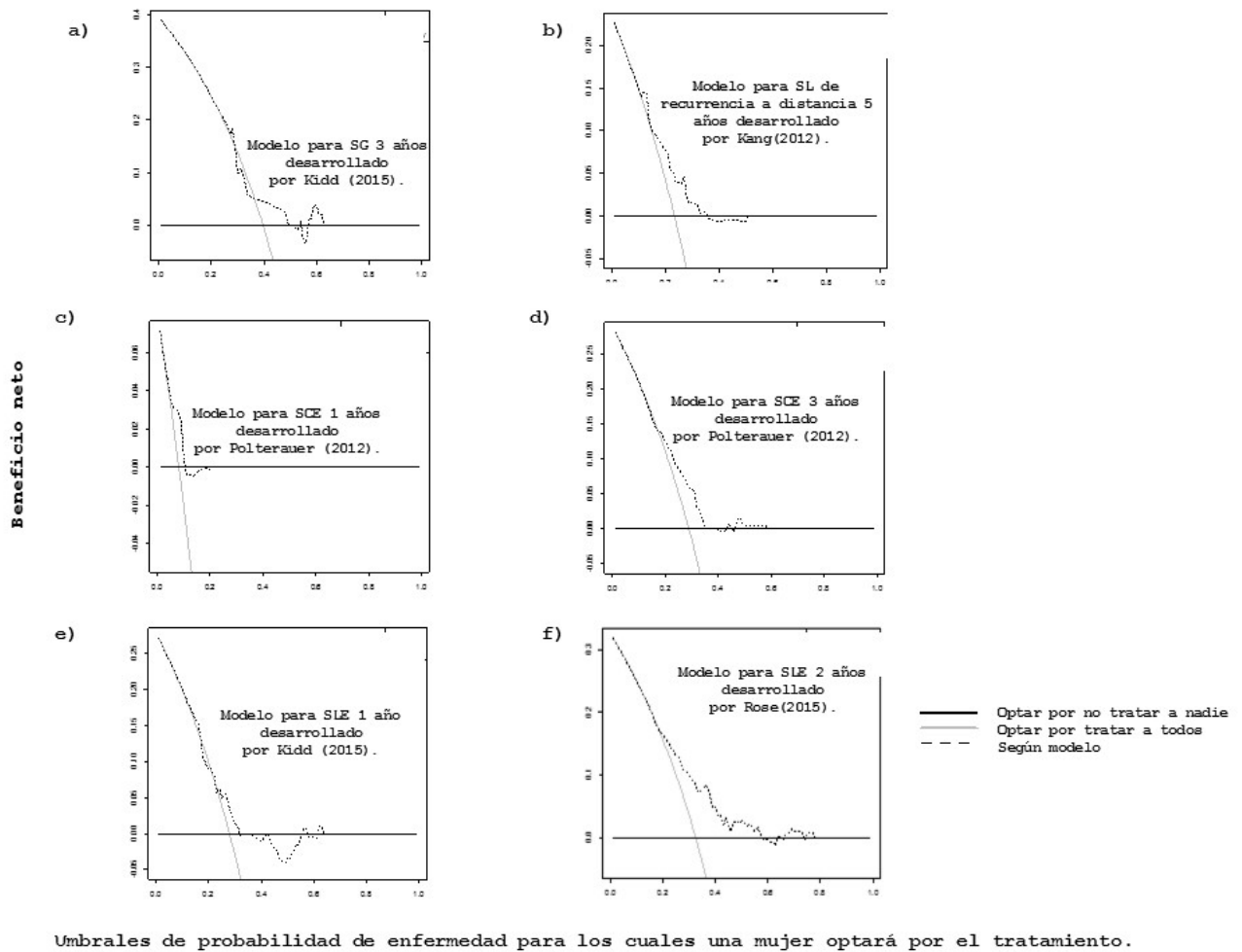


Figura 13. Beneficio neto como función de los umbrales de decisión para distintos modelos de pronóstico.

En la Tabla 15 se muestra el análisis de la curva de decisión para los puntos más importantes. Debido al escaso número de sujetos, menos de 100 sujetos, los datos del análisis de la curva de decisión no se mostraron para los modelos de supervivencia global a 3 años y de supervivencia libre de enfermedad a 1 año del artículo de EA Kidd (2012) (120).

	Modelo	Umbral	Sen	Esp	Vpp	Vpn	Lr+	Lr-	BN todos	BN umbral
E 1y	S Polterauer (2012) (117)	4%	100 (100,100)	0.0 (0.0;0.0)	8 (5;11)	-	1.00 (1.00;1.00)	-	0.52	0.52
		8%	96 (87;100)	28 (22;33)	10 (6;15)	99 (96;100)	1.32 (1.18;1.48)	0.16 (0.02;1.08)	0.01	0.03
		11%	22 (9;39)	85 (81;89)	11 (2;21)	93 (89;96)	1.46 (0.64;3.33)	0.92 (0.74;1.15)	-0.03	0.00
E 3y	S Polterauer (2012) (117)	12%	100 (100,100)	0.0 (0.0;0.0)	29 (23;35)	-	1.00 (1.00;1.00)	-	0.19	0.19
		28.8%	79 (70;89)	41 (34;49)	35 (28;43)	83 (75;91)	1.36 (1.14;1.61)	0.49 (0.29;0.81)	0.06	0.10
		36%	16 (7;25)	88 (83;93)	36 (19;52)	72 (66;78)	1.36 (0.69;2.67)	0.95 (0.85;1.07)	-0.11	0.00
L 2y	PG Rose (2015) (121)	19%	99 (97;100)	3 (1;6)	33 (27;39)	86 (60;100)	1.02 (0.99;1.06)	0.35 (0.04;2.84)	0.16	0.16
		32.9%	82 (74;90)	42 (35;50)	41 (33;48)	83 (76;91)	1.42 (1.21;1.68)	0.41 (0.25;0.68)	0.00	0.08
		58%	25 (16;34)	92 (87;96)	58 (42;74)	72 (66;76)	2.91 (1.59;5.37)	0.82 (0.72;0.94)	-0.61	0.00
Rd 5y	S Kang (2012) (41)	12%	100 (100,100)	0.0 (0.0;0.0)	23 (17;30)	-	1.00 (1.00;1.00)	-	0.13	0.13
		23.3%	53 (37;69)	66 (58;74)	32 (21;44)	82 (74;90)	1.55 (1.05;2.29)	0.72 (0.50;1.02)	0.01	0.05
		36%	11 (1;20)	95 (91;99)	40 (10;70)	78 (71;84)	2.18 (0.65;7.31)	0.94 (0.84;1.06)	-0.19	0.00

Tabla 15. Regiones de validez clínica. E: Supervivencia Causa Específica (causa cáncer); L: Supervivencia Libre de Enfermedad; Rd: Supervivencia libre de Recurrencia a distancia; y: instante (año) de Tiempo de evaluación; Sen: Sensibilidad; Esp: Especificidad; Vpp: Valor predictivo positivo; Vpn: Valor predictivo negativo. Lr: Likelihood ratio (+) positivo y (-) negativo. BN: Beneficio Neto. Sensibilidad, Especificidad y Valor predictivo se muestran en porcentajes.

7.2.4.1 Utilidad clínica del modelo de PG Rose (2015) (121).

Conocemos que la proporción de mujeres diagnosticadas de CCLA con recidiva de la enfermedad a 2 años es entre un 20 y un 30%. Por tanto, la situación es de baja prevalencia. Imaginamos que queremos utilizar el modelo de PG Rose (2015) (121) para incorporar a las pacientes en un ensayo clínico de un nuevo fármaco efectivo. Tres situaciones posibles: el nuevo fármaco es peligroso, el nuevo fármaco es costoso pero no tiene efectos secundarios y el nuevo fármaco es inocuo.

1. En la primera situación, fármaco peligroso, estaremos interesados en reducir al máximo el número de falsos positivos, por tanto, es conveniente que el valor predictivo de un resultado positivo en el análisis de utilidad sea alto (punto C). De esta manera, si determinamos la probabilidad subjetiva de incluir a un sujeto como alta, es decir, necesitamos estar seguros que el paciente va a desarrollar el evento, y se establece un umbral por encima de 60%, la situación llevaría a no incluir a ningún sujeto (Tabla 15).
2. En la situación segunda, el nuevo fármaco es costoso pero no tiene efectos secundarios, es interesante que el punto C (valor predictivo de un resultado positivo) sea alto, pero la probabilidad subjetiva para incluir a un sujeto no necesariamente ha de ser tan alto, por tanto, un umbral de 40%, nos llevaría a incluir o no a las pacientes en función del modelo de pronóstico de PG Rose (2015) (121).
3. En el último caso, nuevo fármaco inocuo, el valor umbral de la probabilidad subjetiva inferior (valor predictivo de un resultado negativo) obtenido en el análisis de la curva de decisión lo interesante es que sea bajo. Debido a que tratar a las pacientes puede evitar la recurrencia y aquellas que no fueran a recurrir de la enfermedad no tendría

ningún efecto sobre ellas el tratamiento. Por tanto, si la probabilidad subjetiva se establece como 5%, se debería tratar a todas las pacientes y un umbral de 20%, llevaría a la utilización del modelo para incluir o no incluir a las pacientes.

8 DISCUSIÓN.

8.1 Resumen de los principales resultados.

Existen 8 estudios en la literatura científica donde se desarrollan un total de 13 modelos con capacidad pronóstica en términos de mortalidad, mortalidad causa específica, recurrencia de la enfermedad, recurrencia a distancia, local o para-aórtica, al momento del diagnóstico en diferentes instantes de tiempo para mujeres diagnosticadas de cáncer de cuello uterino localmente avanzado y tratadas con quimio-radioterapia concurrente.

Los factores pronóstico que usan los modelos de riesgo identificados son la edad, la raza, el estadio FIGO, el tipo histológico, grado de diferenciación, la afectación del ganglio linfático y su localización, volumen del tumor, SUVmax del tumor de cérvix, ECOG, niveles de SCC-Ag y tratamiento recibido.

El resumen de la información detallada en los artículos es aceptable. De la información existente, se desprende la presencia de limitaciones propias por el tipo de estudio, la recogida de la información, el tamaño de la muestra y el análisis de los datos.

De los 11 modelos de pronóstico desarrollados en pacientes donde una parte o todas las mujeres recibieron quimio-radioterapia, 5 modelos (6 momentos en el tiempo) presentan buena validez externa, en términos de discriminación y calibración: el modelo de supervivencia global a 3 años desarrollado por EA Kidd (2012) (120), el modelo de supervivencia causa específica de S Polterauer (2012) para 1 y 3 años, los modelos de supervivencia libre de enfermedad a 1 año de EA Kidd (2012), y a 2 años de PG Rose (2015) y el modelo de recurrencia a distancia de S Kang (2012) a 5 años.

Los 5 modelos de pronóstico con validez estadística presentaron regiones de validez clínica estrechas.

La región de utilidad clínica de un modelo de pronóstico corresponde con la región comprendida entre 1 menos el valor predictivo negativo y el valor predictivo positivo. La clasificación mediante la probabilidad de un modelo de pronóstico no es útil cuando ambos valores corresponde con la prevalencia del evento evaluado.

8.2 Exhaustividad y aplicabilidad de la evidencia.

La RS encontró 8 artículos sobre modelos de pronóstico en pacientes con CCLA. Los 8 artículos conformaban un total de 13 modelos de riesgo. Existen otros modelos de pronóstico desarrollados para mujeres con CCLA con otros tratamientos, la decisión de no validarlos fue tomada debido a que validar externamente en una población tratada con quimiorradioterapia un modelo desarrollado en pacientes con otros tratamientos afectará en las medidas de rendimiento referentes a la calibración, además de la posible existencia de variables en los modelos referentes al tratamiento.

El modelo de SH Shim (2013) (119) para supervivencia causa específica es un modelo sin capacidad discriminativa y con problemas de sobreestimación en la calibración. Este modelo es el único del que se dispuso de los valores originales del nomograma, aun así los problemas detectados en la validez estadística pueden ser debidos a no contemplar alguna variable o variables importantes en el modelo, a la reducción y simplificación del modelos (3 grupos para el tamaño del tumor), o a que existe una variable que se estima de diferente manera. La falta de validez estadística, por tanto, implica que antes de ser llevado a la clínica o usado en ensayos clínicos son necesarios más estudios sobre él.

El modelo de S Kang (2012) (41) para recurrencia a distancia es un modelo con validez estadística y clínica. Siendo población asiática, las mujeres que componen la totalidad de la muestra de desarrollo, se adecuó correctamente en término de validez estadística el modelo a la muestra de validación. No obstante, el decremento en discriminación repercutió negativamente en la amplitud de los rangos de los umbrales de probabilidad. La simplificación del modelo que propone Kang, como era esperable, discrimina peor que el modelo general.

El modelo de JY Tseng (2010) (118) para supervivencia causa específica disminuye su capacidad discriminativa, área ROC = 0.69, de la muestra de desarrollo a la muestra de validación externa, área ROC = 0.64, lo cual es esperable. El modelo falló en calibración, sobreestimando el riesgo de los individuos. Esta sobre valoración del riesgo puede ser debido a no contemplar variables de interés en el modelo, aunque también puede estar influenciado por la simplificación, a través de la categorización, de las variable continuas tamaño del tumor y SCC-Ag. Otros fenómenos que pueden impactar en la mala calibración son la selección de sujetos según el tipo histológico, únicamente células escamosas, que se hizo en el artículo de desarrollo, mientras que en la muestra de validación no se excluye a ninguna paciente según el tipo histológico, así como, la recodificación de la variable ganglio linfático afecto.

Las estratificaciones del riesgo presentadas por JA Liang (2008) (122) para supervivencia libre de enfermedad y recurrencia para-aórtica están basadas en el estadio FIGO y la afectación de los ganglios pélvicos. Estas reducciones simplificadas consiguieron discriminar únicamente en la supervivencia libre de enfermedad.

El modelo de S Polterauer (2012) (117) para supervivencia causa específica es un modelo desarrollado en población europea, teóricamente más semejante a la población española de donde se obtiene la muestra de validación externa. La muestra de desarrollo del modelo seleccionó a toda mujer diagnosticada de cáncer de cuello uterino, no excluyendo a aquellas mujeres con diagnóstico en estadios tempranos. Aun así, el modelo gozó en general de buena validez estadística, aunque perdió calibración a medida que las predicciones fueron a más largo plazo, para los momentos a 1 y 3 años está bien calibrado pero para 5 años no. Esto puede ser debido a que la supervivencia causa específica es un evento competitivo con la muerte por otras causas, y esto influye en la selección de los pacientes a largo plazo. También perdió capacidad discriminativa de la muestra de desarrollo a la muestra de validación, área ROC de 0.72 a 0.63. Esto influyó de manera negativa en la amplitud de los umbrales de probabilidad donde es beneficioso en la práctica clínica, haciendo que los rangos fueran estrechos.

Los 3 modelos de EA Kidd (2012) (120), modelo para supervivencia global, supervivencia causa específica y supervivencia libre de enfermedad, se ven influenciados por la determinación de manera diferente de las variables incluidas en el modelo, técnica de PET en la muestra de desarrollo y cualquier otro método (TC, RM, PET-TC o linfadenectomía paraaórtica) en la muestra de validación externa. Además la cantidad de datos faltantes, con característica no aleatoria, hace que las cualidades de los modelos estén disminuidas. Aun así, son adecuados en términos de discriminación y calibración los modelos de supervivencia global y supervivencia libre de enfermedad. Ambos modelos tienen una particularidad, prescinden del estadio FIGO, variable principal en gran parte de los modelos por estar bastante relacionada con la mortalidad, como se presentó en los diagramas de cajas. El modelo de supervivencia global y supervivencia libre de enfermedad son de utilidad clínica, y

aportan umbrales donde pueden ser de utilidad para el clínico. En el resto de modelos aportado por EA Kidd (2012) (120) no existe evidencia para rechazarlos. En estos modelos los umbrales son más estrechos que los ofrecidos en el modelo de PG Rose (2015) (121) debido al impacto de la baja capacidad discriminativa que poseen.

Por último, el modelo de PG Rose (2015) (121) para supervivencia libre de enfermedad a 2 años fue el que obtuvo mejor discriminación, aunque presenta débil calibración. Llama la atención que en ambos modelos de PG Rose (2015) (121), modelo de supervivencia global y libre de enfermedad, se incrementa la capacidad discriminativa en la muestra de validación externa, área ROC de 0.64 a 0.66 para supervivencia global y de 0.62 a 0.70 para supervivencia libre de enfermedad, con respecto a la que tenían en la muestra de desarrollo. Las muestras de desarrollo y la muestra de validación presentan diferencias en los estadios FIGO y ECOG, menores estadios y mejor ECOG para la muestra de desarrollo, no obstante, debemos tener en cuenta que las pacientes del estudio de PG Rose (2015) (121) provienen de 6 ensayos clínicos (143–148), lo que le aporta mayor similitud entre las pacientes de la muestra de desarrollo, por ser mujeres más seleccionadas y tener criterios de medición más homogéneos, pero mayor diferencias con respecto a muestras externas. Este fenómeno es el que provoca probablemente el incremento en discriminación. Por otro lado, mientras que el modelo de supervivencia libre de enfermedad está calibrado, el modelo de supervivencia global presenta problemas de sobreestimación. La sobreestimación está relacionada con la ausencia de alguna variable importante no incluida en el modelo, podría deberse a la codificación de la variable raza a través de la nacionalidad. Tener la fórmula exacta del modelo permitiría poder llevar a cabo un estudio de puesta a punto, y corregir los coeficientes para que el modelo estuviera bien calibrado y fuera más exacto. Por último, el modelo de

supervivencia libre de enfermedad presenta validez clínica en umbrales de utilidad clínica estrecho, fenómeno explicado debido a su baja discriminación y débil calibración.

8.3 Calidad de la evidencia.

La evaluación de los artículos a través de la guía TRIPOD (47,89) evidenció la carencia de investigación pronóstica de excelencia en el cáncer de cuello uterino localmente avanzado. Ninguno de los artículos cumplió de forma correcta con los 31 ítems de la guía TRIPOD (47,89), bien es cierto, que la guía TRIPOD (47,89) es posterior a todos los artículos. Los potenciales sesgos referidos al tipo de diseño de carácter retrospectivo empleado en la mayoría de los casos, se unen con el inadecuado tamaño de la muestra, la transparencia en cuanto los métodos de análisis o de imputación de datos y las ausencias de definiciones de evento y tiempo hasta el evento. En algunos artículos, se confunde incluso la palabra supervivencia global por supervivencia causa específica. Además, la evaluación de las medidas de rendimiento no fueron descritas en bastantes de los artículos.

Por tanto, debido a la implicación que un pronóstico fallido puede tener en un paciente, de carácter muy perjudicial, se evaluó la calidad global de todos los artículos como aceptable.

8.4 Sesgos potenciales en el proceso.

Previo a la realización de la RS sobre modelos de pronósticos, no hubo una publicación del protocolo siguiendo el estándar del manual Cochrane de revisiones sistemáticas de intervenciones (21), lo que supone una debilidad del estudio. Aunque existieron directrices claras sobre los criterios para que un artículo sobre modelos de pronóstico fuera incluido. Estas directrices fueron usadas en la revisión por pares. Los hallazgos en la revisión mostraron la necesidad de incluir artículos de modelos de pronóstico desarrollados en pacientes con CCLA junto con mujeres con cáncer de cuello uterino en estadios tempranos

(117,138). Esta incorporación en los criterios de elegibilidad es válida y de interés, puesto que el modelo de carácter general ofrece probabilidades de riesgo específica para las mujeres con cáncer de cuello uterino localmente avanzado. Por tanto, aunque se hicieron cambios en las pautas de los estudios a incluir, estos no modificaron el objetivo principal del estudio y de la revisión.

Excepto la construcción del proyecto y la selección definitiva de los modelos de pronóstico en mujeres con cáncer de cuello uterino localmente avanzado que fue llevada a cabo por más de una persona, el resto de tareas, búsqueda y extracción de datos, fue realizado por un único individuo. Aunque, idealmente, las revisiones sistemáticas necesitan ser asumidas por más de una persona (21), lo que asegura la detección de errores y minimiza los posibles sesgos, bien es cierto, que siempre existió el respaldo del grupo como asesores para la resolución de dudas en la revisión, existiendo un consenso posterior en aquellas actividades que se hicieron de manera independiente. Para esta revisión no se contactó con ningún Grupo Cochrane de Revisión, lo que implica una nueva limitación.

La revisión fue orientada a localizar estudios pronósticos y no a evaluar tratamientos o intervenciones, esto creó una búsqueda exhaustiva por distintas razones. Primero, las revisiones sistemáticas requieren una búsqueda amplia, objetiva y reproducible de una gama de fuentes, para identificar tantos estudios relacionados como sea posible (21). En esta RS, se buscó en siete fuentes de información diferentes: Epistemonikos, IBECS (isciii), ISI Web Of Knowledge, Medline (pubmed), EMBASE (Ovid), IME (Índice Médico Español) y LILACS; de esta manera se procuró identificar tantos estudios relacionados como fuera posible, y minimizar los sesgos de selección de aquellos artículos encontrados. Segundo, la identificación de artículos fue complementada con la búsqueda de referencias citadas, lo que permite encontrar artículos enlazados por su cita. Una posible limitación en la ampliación de

fuentes fue la exclusión de bases de datos de literatura gris, tesis y tesinas, búsquedas en la web, y de búsquedas manuales. Se decidió no incluirlas ya que suponían un esfuerzo en tiempo y recurso frente a un esperable e improbable hallazgo. Ese tiempo se asignó al diseño de una estrategia de búsqueda eficiente y a la construcción de un lenguaje óptimo basado en revisiones sistemáticas previas sobre los mismo temas (1,4,8,22,100–105) y búsquedas óptimas construidas (99,109–113). De esta manera, la identificación de artículos resulto fortalecida al combinar palabras de texto libre con términos estandarizados asignados por indexadores (parte de un vocabulario controlado o tesauro) y filtros de búsqueda (99,109–113). Esta estrategia fue utilizada en Medline (pubmed), con el uso de términos MeSH (“Medical Subject Headings”), y en EMBASE (Ovid), donde además de incorporar términos de vocabulario controlado (EMTREE), se añadió a la búsqueda óptima (109,110), operadores de cercanía como “NEAR”. En el resto de fuentes, quitando ISI Web Of Knowledge donde se permiten búsquedas de palabras que tuvieran la misma cadena de letras como raíz, las búsquedas presentaron serias limitaciones al no permitir la consulta de un resumen estructurado para conocer el contenido del artículo y no disponer de un tesauro normalizado (149). Tercero, la pregunta de la revisión se estructuró adaptando la pregunta de las revisiones sobre intervenciones (21), basada en la estructura PICO (Participantes, Intervenciones, Comparaciones y Desenlaces) para la identificación de modelos de pronóstico. Esto, permitió delimitar la pregunta que la revisión quería responder. Así, combinando la estrategia de búsqueda anteriormente descrita, la serie de términos de búsqueda se elaboró siguiendo la situación de salud (cáncer de cérvix) y la finalidad del diseño de estudio (estudios de pronósticos). Esto puede presentar una limitación, porque al no especificar en la búsqueda “cáncer de cérvix localmente avanzado” provocó un aumento de la extensión (o sensibilidad) de la búsqueda, reduciendo su precisión y recuperando más artículos no relevantes, pero por el contrario, permitió capturar artículos de interés que no

mencionaban en el título y ni en el resumen el estado del cáncer (21). De igual manera, la finalidad pronóstica de carácter longitudinal incluyó estudios aleatorizados y no aleatorizados. Esto implicó un incremento en la extensión de la búsqueda por ser más difíciles de encontrar y no existir métodos recomendados para limitar las estrategias de búsqueda para estudios no aleatorizados (21). Finalmente, aunque el objetivo principal del estudio es únicamente validar externamente los modelos de pronóstico existentes para pacientes tratados con quimiorradioterapia, la revisión no se limitó a este tipo de intervención ampliando la búsqueda, incluyendo cualquier modelo desarrollado para mujeres con cáncer de cérvix a excepción de aquellos con datos ómicos.

Hay un alto desarrollo de los aspectos metodológicos que deben cumplir los trabajos con fines pronósticos (149), y en general, existe un acuerdo general en los criterios de verificación para su evaluación metodológica (47,89,102). La guía específica de revisión de la calidad metodológica de los artículos fue la guía TRIPOD (47,89), guía que descompone en aspectos específicos las dimensiones metodológicas propias de un estudio de pronóstico. Una de las fortalezas de esta revisión es el empleo del protocolo TRIPOD (47,89), protocolo que permite evaluar la calidad de la evidencia científica en los artículos obtenidos en la revisión, así como detectar posibles sesgos o errores (21), además de los sabidos como propios según el diseño del estudio (43).

El estudio de validación externa llevado a cabo se realizó bajo un diseño de cohorte retrospectiva. Por tanto, la identificación de mujeres se hizo en una cohorte de mujeres del pasado que describían los criterios de inclusión establecidos en el presente trabajo y, desde ese momento, se siguieron a las pacientes (21). El diseño de estudio planteado para la validación externa se amolda correctamente al carácter longitudinal de los modelos de pronóstico (43,47,89), siendo el diseño de cohorte preferible en la investigación pronóstica a

cualquier otro tipo de diseño, por ejemplo, al estudio de casos y controles o a los estudios de ensayos clínicos (29,43).

La debilidad del diseño de estudio planteado reside en la manera en la que se identificó a las pacientes, es decir, en el carácter retrospectivo del estudio. Si bien es cierto, que es una manera simple, flexible y de bajo coste en la búsqueda y recolección de pacientes (43,47,89), pero que a su vez, provoca gran cantidad de datos perdidos, no recogidos, e información no correcta, con categorías diferentes a las necesarias por el modelo, e incluso medidas con otras técnicas, procedimientos y métodos distintos. Ejemplo de ello fueron, las mediciones para evaluar la afectación de los ganglios linfáticos en los pacientes de la cohorte de validación usando las técnicas de TC, RM, PET-TC o linfadenectomía paraaórtica frente a métodos exclusivos de RM (119) o únicamente de PET (41,120), así como el empleo de nacionalidad para generar la variable raza, o el uso de afectación de los ganglios linfáticos para crear afectación de ganglios pélvicos (41,121,122), paraaórticos (41,119,121) o ganglio linfático afecto mayor (120), o la asignación de un valor en la variable razón entre ganglios linfáticos positivos y negativos (117) por no haber sido recogido en la base de validación. Esta asignación tuvo un impacto directo (no cuantificado en este estudio), provocando posibles sesgos, en la estimación de las probabilidades de supervivencia para cada tiempo específico (124). En cambio, un estudio prospectivo hubiera permitido la medición óptima de los predictores y del evento.

Por último, la información fue recogida por un único centro, y eso implica un menor número de pacientes que conforman la muestra frente a estudios con múltiples centros, donde además las variables son más generales al necesitar consenso y puesta en común entre los diferentes hospitales o centros.

La debilidad más relevante del estudio es el número de pacientes incluidos en la validación externa (26,27,47,89,114,115). Las recomendaciones en la literatura del tamaño de la muestra para desarrollar estudios de validación externas están basadas en estimar el número de pacientes necesarios basado en contraste de hipótesis (27) o en la examinación del comportamiento de las medidas de resumen para producir exactas y precisas estimaciones que reflejen la verdadera población subyacente (26,115). Estas recomendaciones (26,27,114,115) coinciden en que 100 eventos, incluso 200, es el número mínimo aproximado para construir muestras de validación externa. En nuestro caso, el número de pacientes fue determinado en función de los pacientes incluidos en la base de datos asistencial (47,89), 346 pacientes, de las cuales conformaron el estudio 297 mujeres. El número de eventos para todos los estudios considerados fue inferior a 100, llegando a ser dramático en los modelos de riesgo evaluados a un año de S Polterauer (2012) (117) y de EA Kidd (2012) (120), esto implica que es probable que el estudio de validación sea poco fiable, impreciso y esté sesgado (26). Para compensar la falta de tamaño muestral teórico, los intervalos de confianza fueron aportados para mostrar el conjunto de verdaderas pero desconocidas valor de las medidas de rendimiento de los modelos de pronóstico que son estadísticamente compatibles con las medidas de rendimiento observadas (127).

Un paso necesario e importante para la realización del estudio de validación externa es disponer de la fórmula del modelo original, y con ella calcular las probabilidades predichas para cada individuo (73). Los modelos objetivo de esta revisión, modelos tiempo dependientes, fueron desarrollados casi en su totalidad a través de modelos de riesgos proporcionales de Cox, lo que implica que es necesario para validarlos disponer de los coeficientes del modelo (o hazard ratio) junto con la función de supervivencia o la función de riesgo basal (88). Fue frecuente encontrar únicamente la herramienta gráfica o nomogramas

(39,40) sin información adicional o incompleta (únicamente los hazard ratio de los modelos multivariantes), y por tanto, fue necesario contactar con los autores principales de los estudios solicitando la información del modelo o la correspondencia exacta del nomograma con los ejes (21). Únicamente el autor del modelo para supervivencia causa específica a 3 y 5 años (119) aportó los valores exactos para el nomograma. Y por tanto, junto con las probabilidades predichas por los modelos implementados vía web (41,117) fueron los únicos que las probabilidades individuales fueron exactas con las del modelo desarrollado (47,89). Nuevamente, aunque el valor de probabilidad fue calculado a través de una función programada en R (123), para evitar diferencias entre perfiles iguales según el momento en el que se proceda al cálculo, no utilizar la función o valores originales provocó errores de transcripción y de redondeo (100), generando ruido, y posibles sesgos en las medidas de evaluación de las capacidades predictivas de los modelos, siendo una gran limitación del estudio de validación externa.

Esta falta de traspaso de información entre investigadores limitó el estudio de relación entre las pacientes de la muestra de desarrollo y la muestra de validación externa (43,48,150) a la comparación de las diferencias entre cada uno de las variables biográficas, biológicas, terapéuticas y clínicas. Esta aproximación es útil para comparar características específicas a lo largo de las muestras de estudio, pero extraer un juicio global de la homogeneidad o relación entre las muestras resulta complicado (48). Añadir que la información de las características de los sujetos de los artículos originales, en caso de que estén presentes, se obtiene de forma resumida y en diferentes formatos (p.ej. las edades como medias o medianas, con las desviaciones estándar o los rangos; el sexo como porcentajes o recuentos; y algunas de ellas para todo el estudio o para cada grupo de estadio FIGO por separado), lo que añade variabilidad a la forma de presentar las características de las muestras.

Por último, exceptuando la asignación de un valor promedio en la variable razón entre ganglios linfáticos positivos y negativos para el modelo de supervivencia causa específica (mortalidad debida al cáncer) de S Polterauer (2012) (117), no se hizo ninguna imputación de datos. Es decir, únicamente se trabajó con los datos disponibles. Esto supone, una limitación del estudio (47,89,100), porque reduce la información disponible. En nuestro caso, el único modelo afectado fue el modelo de EA Kidd (2012) (120), que presenta menos de 40 eventos en los modelos y menos de 100 sujetos, donde el origen de las pérdidas es de carácter no aleatorio, debido a que está asociadas con la técnica del PET. Se descartó la imputación, debido a que se debía incorporar más del 50% de los datos. En el resto de modelos las pérdidas únicamente son de 2 a 1 individuo. En ningún caso, se reemplazó la información debido a que las medidas de rendimiento global tampoco iban a sufrir variaciones.

8.5 Acuerdos y desacuerdos con otros estudios o revisiones.

En la literatura científica, existen diferentes trabajos donde se han revisado modelos de pronóstico (100,102,149), existiendo trabajos centrados sobre modelos de riesgo en cáncer (103–105), y para determinados cánceres, por ejemplo la revisión sobre modelos de pronóstico en el cáncer de mamá (105). La mala calidad general de los modelos de pronóstico desarrollados, al igual que lo encontrado en esta revisión, ha hecho que se dé un paso más en la investigación pronostica (25,30,83,85), desarrollando guías para la evaluación resumida en los artículos (47,89,90,100,149). En esta tesis, se eligió la guía TRIPOD (47,89), por ser una guía específica para modelos de pronósticos, y se evaluó de la misma manera que en el artículo de A Gómez-de-la-Cámara (2010) (149), donde se utilizó una guía general para estudios observacionales.

El trabajo presentado por I Stegeman (2012) (52) es una RS sobre tres tipos de cánceres para identificar modelos de pronóstico en población sin síntomas en donde existen programas de cribado: cáncer de mamá, cervix y colón. En la presente tesis, la población de referencia es población diagnosticada de cáncer de cuello uterino localmente avanzado. La revisión de I Stegeman (2012) (52) presenta los elementos y métodos que permiten hacer una búsqueda sistemática y explícita con el fin de minimizar sesgos y aportar resultados más fiables. En ella aparecen reflejado el tipo de diseño, sujeto y la definición de modelo de pronóstico, así como las base de datos consultadas (Medline y Embase), la estrategia de búsqueda, las palabras de texto comunes y términos estandarizados asignados por indexadores. Además se detalla cómo se hace la identificación de los artículos, y de presentar la información de cada artículo. La estructura de la revisión es similar a la que se propone en el manual Cochrane de revisiones sistemáticas de intervenciones (21) , manual de referencia en la presente tesis. Ambas revisiones coincidimos en lo aceptable, en la tesis además evaluado con la guía TRIPOD (47,89), pero no excelente resumen de la información de los artículos encontrados y resumimos la información centrándonos en aspectos como el diseño del estudio, los criterios de inclusión y exclusión, el tamaño de la muestra, las variables que conforman los modelos y las medidas de rendimiento pronóstico (125). En el trabajo I Stegeman (2012) (52) no se presenta ninguna validación externa de los modelos evaluados.

La RS presentada por P Biewenga (2009) (56) se desarrolló con el objetivo de validar externamente modelos de pronósticos para predecir supervivencia y recurrencia en mujeres con estadios tempranos de cáncer de cervix. Por tanto, las poblaciones de los estudios son diferentes, estadios tempranos (56) frente a estadios localmente avanzados de la tesis. En esta revisión, se detalla de manera sucinta los pasos para llevar a cabo la revisión e identificación de los artículos, únicamente bases de datos consultadas, años, lenguaje de la publicación y

proceso de inclusión y exclusión. No se especifica la estrategia de búsqueda a diferencia de la presente tesis (11.1. Estrategias de búsqueda.). Los dos trabajos presentan la información de manera homogénea, en cuanto a características basales de los pacientes, tratamiento principal y estadios FIGO entre otros, así como los eventos a identificar y el carácter retrospectivo de la validación externa. Hacemos notar que debido al tamaño de la muestra de 563 mujeres con probabilidad de supervivencia causa específica de 91% y libre de enfermedad de 93% a 5 años, el número de eventos es inferior a 100, de tal manera que la validación externa es cuestionable al igual que ocurre en la presentada en la tesis (26,27,114,115). Y la diferencia radical entre las dos validaciones externas, es que mientras en la tesis se presenta medidas de rendimiento (31,43,47,89,115,125) de los modelos de pronóstico en cuanto a discriminación y calibración, en el trabajo de P Biewenga (2009) (56) se especifica calibración como único objetivo. Los modelos presentados (35,57–67), no son evaluados en cuanto a su utilidad clínica (72,80) frente a lo mostrado en la tesis. Al igual que en el trabajo de P Biewenga (2009) (56), no todos los modelos de pronóstico encontrados fueron validados externamente, en este caso, debido a que el tratamiento de quimioradioterapia mejora la supervivencia general y la supervivencia libre de progresión (21), además de la recurrencia local y a distancia. Validar externamente en una población tratada con quimioradioterapia un modelo desarrollado en pacientes con otros tratamientos afectaría teóricamente a las medidas de rendimiento referentes a la calibración.

En la literatura no se encuentran más revisiones sistemáticas en búsqueda de modelos de pronóstico en el cáncer de cérvix. Sí se encontraron artículos en los que se valida externamente modelos de pronóstico para el cáncer de cuello uterino, por ejemplo, el modelo de DY Kim (2014) (38), para mujeres con cáncer de cuello uterino en estadios tempranos, modelo que permite estimar la probabilidad de tener metástasis en los ganglios linfáticos con

información preoperatoria, a través de una herramienta gráfica o nomograma (39,40). Este modelo de riesgo (38) es desarrollado y de manera consecutiva validado externamente por los mismos investigadores (46) que lo conformaron. De igual manera ocurre, para el modelo pronóstico de S Kang (2012) (41) para predecir recurrencia a distancia en pacientes con cáncer de cérvix avanzado y tratados con quimiorradioterapia. Estas validaciones externas aportan menos evidencia (38) que las realizadas por otros investigadores distintos a los que lo generaron como es el caso de esta tesis. Ambos modelos, también presentan un número inferior a 100 de eventos (26,27,114,115), 6 de 304 para la muestra de desarrollo y 38 de 189 para validación para el modelo de DY Kim (2014) (38), y 92 de 748 para la muestra de desarrollo y 32 de 167 para validación en el modelo de S Kang (2012) (41), al igual que ocurre en esta tesis. Ambos modelos estimaron los medidas de calibración, gráfico de calibración o test de Hosmer-Lemeshow, y de discriminación, estadístico C igual a 0.878 para la muestra de desarrollo y 0.825 en la muestra de validación en el modelo de DY Kim (2014) (38), y de 0.70 a 0.73 en la muestra de validación para el modelo de S Kang (2012) (41). Generalmente, las medidas de rendimiento decrecen cuando el modelo es aplicado a una muestra de validación. A pesar de eso, en este caso se incrementa, explicado probablemente por la heterogeneidad de las muestras (150). Este mismo fenómeno ocurre con los modelos de PG Rose (2015) (121) en la muestra de validación del Servicio de Oncología Radioterápica del Hospital Universitario 12 de Octubre, para ambos modelos el valor puntual del área bajo la curva se incrementa, de 0.64 a 0.66 para el modelo de supervivencia global, y de 0.62 a 0.70 para el modelo libre de enfermedad, probablemente debido a que las mujeres del estudio de PG Rose (2015) (121) pertenecen a pacientes de 6 ensayos clínicos (143–148), lo que hace que las enfermas sean más homogéneas que las incluidas en la muestra del Hospital Universitario 12 de Octubre. También en CCLA encontramos ejemplos de validación externa de carácter prospectivo (15) sobre un modelo pronóstico desarrollado

(151), no ofreciendo los problemas con el tamaño muestral que los anteriores trabajos, pero sin presentar datos de discriminación y calibración.

8.6 Implicaciones para la práctica.

Un modelo para pacientes con cáncer de cuello uterino localmente avanzado incluido en la práctica habitual ayudaría al médico a tomar decisiones sobre la mujer con un nuevo diagnóstico de cáncer de cérvix (32). La identificación de mujeres con mayor riesgo, o con un posible mayor beneficio nos permitiría avanzar en la terapia dirigida (25), identificando aquellas mujeres donde el tratamiento es beneficioso y debería ser usado, aquellas en las que es inefectivo y se debería complementar con otras estrategias, y donde el beneficio es suficiente para algunos sujetos, pero no para todos y el tratamiento adicional debería ser consensuado con el paciente (72). La utilidad clínica de los modelos pronósticos es beneficiosa, además, para informar al paciente y su familia del curso de la enfermedad, sobre las pautas a seguir, así como, para selección de pacientes para inclusión en ensayos clínicos (74). Ejemplo: evaluación de toxicidad tardía de la quimiorradiación concomitante en pacientes con buen pronóstico de supervivencia global o repercusión quimioterapia adyuvante después de la quimiorradiación concomitante en la supervivencia global después de seleccionar a pacientes con riesgo de recaída de la enfermedad.

8.7 Implicaciones para la investigación.

Debido a que los modelos de pronóstico desarrollados para mujeres diagnosticadas de cáncer de cuello uterino localmente avanzado muestran débil validez estadística y clínica (32), parece adecuado llevar a cabo nuevos estudios prospectivos donde se validen los modelos empleando las fórmulas exactas que los determinan, y no herramientas gráficas o nomogramas (39,40). La colaboración de los investigadores que desarrollaron los modelos

parece fundamental en este proceso de validez externa prospectivo, del que podrían obtenerse valores de utilidad clínica (72,74), así como, desarrollarse a partir de ellos, estudios de evaluación del impacto, donde se evaluaría si realmente el modelo aporta información en la práctica clínica, y realmente cambia la conducta del médico. La evidencia extraída en el proceso de investigación pronóstica (25,30,83,85) nos permitiría incluir el modelo dentro de la práctica clínica real y en las guías de recomendaciones de la práctica clínica (30).

9 CONCLUSIONES.

Las variables con relevancia pronóstica son la edad, la raza, el estadio FIGO, el tipo histológico, grado de diferenciación, la afectación del ganglio linfático y su localización, volumen del tumor, SUVmax del tumor de cérvix, ECOG, niveles de SCC-Ag y tratamiento recibido.

Los modelos de supervivencia global a 3 años desarrollado por EA Kidd (2012) (120), el modelo de supervivencia causa específica de S Polterauer (2012) (117) para 1 y 3 años, los modelos de supervivencia libre de enfermedad a 1 año de EA Kidd (2012) (120), y a 2 años de PG Rose (2015) (121) y el modelo de recurrencia a distancia de S Kang (2012) (41) a 5 años, son válidos en una muestra externa respecto a donde fueron desarrollados.

No se encontró evidencia suficiente para descartar la validez externa del modelo de EA Kidd (2012) para supervivencia causa específica en los dos momentos del tiempo evaluados.

Las regiones de utilidad clínica de los modelos de pronóstico desarrollados para mujeres diagnosticadas de CCLA para predecir un evento tiempo son estrechos.

El resumen de la información aportado en los artículos sobre desarrollo de modelos de pronóstico en el marco del CCLA es mejorable.

Los modelos de pronóstico desarrollados para mujeres diagnosticadas de CCLA para predecir un evento tiempo dependiente tienen en su mayoría una moderada validez estadística.

Bajo el análisis de la curva de decisión, el umbral o probabilidad a priori para definir grupos de riesgo bajo, moderado o alto para recaída debe ser establecido en función de la implicación que tenga la consideración de positivo para la enfermedad y errar un pronóstico.

La adición de al menos uno o más puntos de corte informativos en el gráfico del área bajo la curva o descrita en el texto sobre la probabilidad de un modelo de pronóstico permite caracterizar mejor la prueba así como evaluar la capacidad discriminativa de la prueba.

La región de utilidad clínica de un modelo de pronóstico en el ámbito del análisis de la curva de decisión está comprendida entre 1 menos el valor predictivo negativo y el valor predictivo positivo.

En ausencia de un modelo de pronóstico, la estrategia de clasificar a todos los sujetos como positivos para el desenlace tiene el mismo beneficio neto que la estrategia de clasificar a todos como negativos cuando, el umbral de decisión coincide con la proporción de desenlace de la muestra de validación.

10 BIBLIOGRAFÍA.

1. Waggoner SE. Cervical cancer. *Lancet Lond Engl*. 28 de junio de 2003;361(9376):2217-25.
2. Organization WH, others. International Agency for Research in Cancer. *Globocan 2012: estimated cancer incidence, mortality and prevalence worldwide 2012*. 2013.
3. Extramural Committee to Assess Measures of Progress Against Cancer. Measurement of progress against cancer. *J Natl Cancer Inst*. 16 de mayo de 1990;82(10):825-35.
4. Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. *Lancet Lond Engl*. 7 de septiembre de 2013;382(9895):889-99.
5. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471–1474.
6. Quiero T, Puñal J. Métodos automatizados de lectura de citología cervical uterina. *Desarro Act Red Esp Agencias Eval Tecnol Prestac SNS*. 2013;
7. Salgado Á, Queiro T, Sobrido M, Cerdà T. Revisión sistemática de la evidencia científica: nuevos métodos para el cribado de cáncer de cérvix. En: Castells X, Sala M, Ascunce N, Salas D, Zubizarreta R, Casamitjana M, editores. *Descripción del cribado del cáncer en España Proyecto DESCRIC*. Madrid: Plan de Calidad para el Sistema Nacional de Salud. Ministerio de Sanidad y Consumo. Agència d’Avaluació de Tecnologia i Recerca Mèdiques de Catalunya; 2007. p. 225-74.
8. Collaboration C for CCM-A, others. Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: a systematic review and meta-analysis of individual patient data from 18 randomized trials. *J Clin Oncol*. 2008;26(35):5802–5812.
9. Straughn J, Yashar C. Management of locally advanced cervical cancer. En: Goff B, Mundt A, Dizon D, Vora S, Falk S, editores. *UptoDate [Internet]*. 2015 [citado 5 de febrero de 2016]. Disponible en: http://www.uptodate.com/contents/management-of-locally-advanced-cervical-cancer?source=see_link
10. Green J, Kirwan J, Tierney J, Vale C, Symonds P, Fresco L, et al. Concomitant chemotherapy and radiation therapy for cancer of the uterine cervix. *Cochrane Database Syst Rev*. 2005;(3):CD002225.
11. Committee on Practice Bulletins—Gynecology. ACOG Practice Bulletin Number 131: Screening for cervical cancer. *Obstet Gynecol*. noviembre de 2012;120(5):1222-38.
12. Jin J. JAMA patient page. Screening for cervical cancer. *JAMA*. 3 de diciembre de 2014;312(21):2302.
13. Volerman A, Cifu AS. Cervical cancer screening. *JAMA*. 3 de diciembre de 2014;312(21):2279-80.

14. PDQ Cancer Information Summaries. National Cancer Institute (US); 2002.
15. Tewari KS, Sill MW, Monk BJ, Penson RT, Long HJ, Poveda A, et al. Prospective Validation of Pooled Prognostic Factors in Women with Advanced Cervical Cancer Treated with Chemotherapy with/without Bevacizumab: NRG Oncology/GOG Study. *Clin Cancer Res Off J Am Assoc Cancer Res*. 15 de diciembre de 2015;21(24):5480-7.
16. Barbera L, Thomas G. Management of early and locally advanced cervical cancer. *Semin Oncol*. abril de 2009;36(2):155-69.
17. Straughn J, Yashar C. Management of early-stage cervical cancer. En: Goff B, Mundt A, Dizon D, Vora S, Falk S, editores. UptoDate [Internet]. 2015 [citado 5 de febrero de 2016]. Disponible en: http://www.uptodate.com/contents/management-of-early-stage-cervical-cancer?source=see_link
18. National Cancer Institute. NCI Issues Clinical Announcement on Cervical Cancer: Chemotherapy Plus Radiation Improves Survival [Internet]. 1999 [citado 5 de febrero de 2016]. Disponible en: <http://nihprod.cit.nih.gov/news/pr/feb99/nci-22.htm>
19. Thomas GM. Improved Treatment for Cervical Cancer — Concurrent Chemotherapy and Radiotherapy. *N Engl J Med*. 15 de abril de 1999;340(15):1198-200.
20. Gill S, Loprinzi CL, Sargent DJ, Thomé SD, Alberts SR, Haller DG, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol Off J Am Soc Clin Oncol*. 15 de mayo de 2004;22(10):1797-806.
21. Higgins JP., Green S, Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions [Internet]. 2011 [citado 5 de febrero de 2016]. Disponible en: <http://handbook.cochrane.org>
22. Collaboration C for CCM, others. Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: individual patient data meta-analysis. *Cochrane Database Syst Rev*. 2010;(1):CD008285.
23. Tangjitgamol S, Katanyoo K, Laopaiboon M, Lumbiganon P, Manusirivithaya S, Supawattanabodee B. Adjuvant chemotherapy after concurrent chemoradiation for locally advanced cervical cancer. *Cochrane Database Syst Rev*. 2014;12:CD010401.
24. Rydzewska L, Tierney J, Vale CL, Symonds PR. Neoadjuvant chemotherapy plus surgery versus surgery for cervical cancer. *Cochrane Database Syst Rev*. 2012;12:CD007406.
25. Hingorani AD, Windt DA van der, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 4: stratified medicine research. *BMJ*. 2013;346:e5793.
26. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 30 de enero de 2016;35(2):214-26.

27. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol.* mayo de 2005;58(5):475-83.
28. Baker SG, Kramer BS, Corle D. The fallacy of enrolling only high-risk subjects in cancer prevention trials: Is there a «free lunch»? *BMC Med Res Methodol.* 2004;4:24.
29. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375.
30. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10(2):e1001381.
31. Harrell F. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis.* Springer; 2015.
32. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 29 de febrero de 2000;19(4):453-73.
33. Núñez E, Steyerberg EW, Núñez J. [Regression modeling strategies]. *Rev Esp Cardiol.* junio de 2011;64(6):501-7.
34. Delgado G. Stage IB squamous cancer of the cervix: the choice of treatment. *Obstet Gynecol Surv.* marzo de 1978;33(3):174-83.
35. Sedlis A, Bundy BN, Rotman MZ, Lentz SS, Muderspach LI, Zaino RJ. A randomized trial of pelvic radiation therapy versus no further therapy in selected patients with stage IB carcinoma of the cervix after radical hysterectomy and pelvic lymphadenectomy: A Gynecologic Oncology Group Study. *Gynecol Oncol.* mayo de 1999;73(2):177-83.
36. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA.* 12 de septiembre de 2007;298(10):1209-12.
37. Kattan MW, Vickers AJ. Incorporating predictions of individual patient risk in clinical trials. *Urol Oncol Semin Orig Investig.* julio de 2004;22(4):348-52.
38. Kim D-Y, Shim S-H, Kim S-O, Lee S-W, Park J-Y, Suh D-S, et al. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. *Br J Cancer.* 7 de enero de 2014;110(1):34-41.
39. Kattan MW, Marasco J. What is a real nomogram? *Semin Oncol.* febrero de 2010;37(1):23-6.
40. Vickers AJ, Cronin AM. Everything you always wanted to know about evaluating prediction models (but were too afraid to ask). *Urology.* diciembre de 2010;76(6):1298-301.
41. Kang S, Nam B-H, Park J-Y, Seo S-S, Ryu S-Y, Kim JW, et al. Risk assessment tool for distant recurrence after platinum-based concurrent chemoradiation in patients with

- locally advanced cervical cancer: a Korean gynecologic oncology group study. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 de julio de 2012;30(19):2369-74.
42. Green JA, Kirwan JM, Tierney JF, Symonds P, Fresco L, Collingwood M, et al. Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: a systematic review and meta-analysis. *Lancet Lond Engl*. 8 de septiembre de 2001;358(9284):781-6.
 43. Steyerberg E. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer Science & Business Media; 2008.
 44. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 28 de febrero de 1996;15(4):361-87.
 45. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338:b604.
 46. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 16 de marzo de 1999;130(6):515-24.
 47. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Diabet Med J Br Diabet Assoc*. febrero de 2015;32(2):146-54.
 48. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. marzo de 2015;68(3):279-89.
 49. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart Br Card Soc*. mayo de 2012;98(9):691-8.
 50. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. enero de 2008;61(1):76-86.
 51. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
 52. Stegeman I, Bossuyt PM. Cancer risk models and preselection for screening. *Cancer Epidemiol*. octubre de 2012;36(5):461-9.
 53. Sengupta S, Chaudhuri SK, Das P, Raichaudhuri B, Pramanik R. Integrated approach to prediction of cervical intra-epithelial neoplasia status through a case-control study. *Int J Cancer J Int Cancer*. 19 de febrero de 1999;84(1):69-73.
 54. Wilkinson CE, Peters TJ, Stott NC, Harvey IM. Prospective evaluation of a risk scoring system for cervical neoplasia in primary care. *Br J Gen Pract J R Coll Gen Pract*. agosto de 1994;44(385):341-4.

55. Patil V, Wahab SN, Zodpey S, Vasudeo ND. Development and validation of risk scoring system for prediction of cancer cervix. *Indian J Public Health*. marzo de 2006;50(1):38-42.
56. Biewenga P, van der Velden J, Mol BWJ, Stalpers LJA, Schilthuis MS, van der Steeg JW, et al. Validation of existing prognostic models in patients with early-stage cervical cancer. *Gynecol Oncol*. noviembre de 2009;115(2):277-84.
57. Delgado G, Bundy B, Zaino R, Sevin BU, Creasman WT, Major F. Prospective surgical-pathological study of disease-free interval in patients with stage IB squamous cell carcinoma of the cervix: a Gynecologic Oncology Group study. *Gynecol Oncol*. septiembre de 1990;38(3):352-7.
58. Kamura T, Tsukamoto N, Tsuruchi N, Saito T, Matsuyama T, Akazawa K, et al. Multivariate analysis of the histopathologic prognostic factors of cervical cancer in patients undergoing radical hysterectomy. *Cancer*. 1 de enero de 1992;69(1):181-6.
59. Sevin BU, Lu Y, Bloch DA, Nadji M, Koechli OR, Averette HE. Surgically defined prognostic parameters in patients with early cervical carcinoma. A multivariate survival tree analysis. *Cancer*. 1 de octubre de 1996;78(7):1438-46.
60. Yuan CC, Wang PH, Lai CR, Yen MS, Chen CY, Juang CM. Prognosis-predicting system based on factors related to survival of cervical carcinoma. *Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet*. noviembre de 1998;63(2):163-7.
61. Kristensen GB, Abeler VM, Risberg B, Trop C, Bryne M. Tumor size, depth of invasion, and grading of the invasive tumor front are the main prognostic factors in early squamous cell cervical carcinoma. *Gynecol Oncol*. agosto de 1999;74(2):245-51.
62. Grisar DA, Covens A, Franssen E, Chapman W, Shaw P, Colgan T, et al. Histopathologic score predicts recurrence free survival after radical surgery in patients with stage IA2-IB1-2 cervical carcinoma. *Cancer*. 15 de abril de 2003;97(8):1904-8.
63. Ho C-M, Chien T-Y, Huang S-H, Wu C-J, Shih B-Y, Chang S-C. Multivariate analysis of the prognostic factors and outcomes in early cervical cancer patients undergoing radical hysterectomy. *Gynecol Oncol*. mayo de 2004;93(2):458-64.
64. Lai C-H, Chang C-J, Huang H-J, Hsueh S, Chao A, Yang J-E, et al. Role of human papillomavirus genotype in prognosis of early-stage cervical cancer undergoing primary surgery. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 de agosto de 2007;25(24):3628-34.
65. Van de Putte G, Lie AK, Vach W, Baekelandt M, Kristensen GB. Risk grouping in stage IB squamous cell cervical carcinoma. *Gynecol Oncol*. octubre de 2005;99(1):106-12.
66. Alvarez RD, Soong SJ, Kinney WK, Reid GC, Schray MF, Podratz KC, et al. Identification of prognostic factors and risk groups in patients found to have nodal metastasis at the time of radical hysterectomy for early-stage squamous carcinoma of the cervix. *Gynecol Oncol*. noviembre de 1989;35(2):130-5.

67. Sartori E, Tisi G, Chiudinelli F, La Face B, Franzini R, Pecorelli S. Early stage cervical cancer: adjuvant treatment in negative lymph node cases. *Gynecol Oncol.* octubre de 2007;107(1 Suppl 1):S170-174.
68. Mize BB, Moron A, Papavlassopoulos A, Tolentino J, Talib M, Salame GM, et al. External validation of nomograms predicting survival of women with locally advanced cervical cancer. *Gynecol Oncol.* junio de 2016;141:153.
69. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Mak Int J Soc Med Decis Mak.* febrero de 2015;35(2):162-9.
70. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* 2008;8:53.
71. Weinstein MC, Fineberg HV. *Clinical Decision Analysis.* 1980. 351 p.
72. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak Int J Soc Med Decis Mak.* diciembre de 2006;26(6):565-74.
73. Van Calster B, Steyerberg EW, Harrell FH. Risk Prediction for Individuals. *JAMA.* 3 de noviembre de 2015;314(17):1875.
74. Vickers AJ, Kramer BS, Baker SG. Selecting patients for randomized trials: a systematic approach based on risk group. *Trials.* 2006;7:30.
75. Vickers AJ, Kattan MW, Daniel S. Method for evaluating prediction models that apply the results of randomized trials to individual patients. *Trials.* 2007;8:14.
76. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ.* 18 de noviembre de 1995;311(7016):1356-9.
77. Pauker SG, Kassirer JP. Therapeutic Decision Making: A Cost-Benefit Analysis. *N Engl J Med.* 31 de julio de 1975;293(5):229-34.
78. Pauker SG, Kassirer JP. The Threshold Approach to Clinical Decision Making. *N Engl J Med.* 15 de mayo de 1980;302(20):1109-17.
79. Steyerberg EW, Vickers AJ. Decision curve analysis: a discussion. *Med Decis Mak Int J Soc Med Decis Mak.* febrero de 2008;28(1):146-9.
80. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA.* 27 de enero de 2015;313(4):409-10.
81. de la Cámara AG. Caracterización de pruebas diagnósticas. *Medicine (Baltimore).* 1998;7(104):4872–4877.
82. Peirce CS. The numerical measure of the success of predictions. *Science.* 14 de noviembre de 1884;ns-4(93):453-4.

83. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis Research Strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595.
84. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart Br Card Soc*. mayo de 2012;98(9):683-90.
85. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380.
86. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. septiembre de 2003;56(9):826-32.
87. Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, Lee KL, Habbema JD. Prognostic models based on literature and individual patient data in logistic regression analysis. *Stat Med*. 30 de enero de 2000;19(2):141-60.
88. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13:33.
89. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 6 de enero de 2015;162(1):W1-73.
90. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. octubre de 2014;11(10):e1001744.
91. Shariat SF, Karakiewicz PI, Roehrborn CG, Kattan MW. An updated catalog of prostate cancer predictive tools. *Cancer*. 1 de diciembre de 2008;113(11):3075-99.
92. van der Leeuw J, van Dieren S, Beulens JWJ, Boeing H, Spijkerman AMW, van der Graaf Y, et al. The validation of cardiovascular risk scores for patients with type 2 diabetes mellitus. *Heart Br Card Soc*. febrero de 2015;101(3):222-9.
93. Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AMW, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ*. 2012;345:e5900.
94. Steyerberg EW, Homs MYV, Stokvis A, Essink-Bot M-L, Siersema PD, SIREC Study Group. Stent placement or brachytherapy for palliation of dysphagia from esophageal cancer: a prognostic model to guide treatment selection. *Gastrointest Endosc*. septiembre de 2005;62(3):333-40.

95. Breems DA, Van Putten WLJ, Huijgens PC, Ossenkoppele GJ, Verhoef GEG, Verdonck LF, et al. Prognostic index for adult patients with acute myeloid leukemia in first relapse. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 de marzo de 2005;23(9):1969-78.
96. Kattan MW, Gönen M, Jarnagin WR, DeMatteo R, D'Angelica M, Weiser M, et al. A nomogram for predicting disease-specific survival after hepatic resection for metastatic colorectal cancer. *Ann Surg*. febrero de 2008;247(2):282-7.
97. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA*. 14 de febrero de 2007;297(6):611-9.
98. Genders TSS, Steyerberg EW, Alkadhi H, Leschka S, Desbiolles L, Nieman K, et al. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. *Eur Heart J*. junio de 2011;32(11):1316-30.
99. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeftang M, Moons KGM, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PloS One*. 2012;7(2):e32844.
100. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40.
101. Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. *J Urol*. mayo de 2001;165(5):1562-8.
102. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med*. 21 de marzo de 2006;144(6):427-37.
103. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC Med*. 2010;8:21.
104. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med*. 2010;8:20.
105. Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest*. marzo de 2009;27(3):235-43.
106. Pérez-Regadera J, Sánchez-Muñoz A, De-la-Cruz J, Ballestín C, Lora D, García-Martín R, et al. Cisplatin-based radiochemotherapy improves the negative prognosis of c-erbB-2 overexpressing advanced cervical cancer. *Int J Gynecol Cancer Off J Int Gynecol Cancer Soc*. enero de 2010;20(1):164-72.
107. Pérez-Regadera J, Sánchez-Muñoz A, De-la-Cruz J, Ballestín C, Lora D, García-Martín R, et al. Impact of epidermal growth factor receptor expression on disease-free survival and rate of pelvic relapse in patients with advanced cancer of the cervix treated with chemoradiotherapy. *Am J Clin Oncol*. agosto de 2011;34(4):395-400.

108. Pérez-Regadera J, Sánchez-Muñoz A, De-la-Cruz J, Ballestín C, Lora D, García-Martín R, et al. Negative prognostic impact of the coexpression of epidermal growth factor receptor and c-erbB-2 in locally advanced cervical cancer. *Oncology*. 2009;76(2):133-41.
109. Holland JL, Wilczynski NL, Haynes RB, Hedges Team. Optimal search strategies for identifying sound clinical prediction studies in EMBASE. *BMC Med Inform Decis Mak*. 2005;5:11.
110. Wilczynski NL, Haynes RB. Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. *J Am Med Inform Assoc JAMIA*. agosto de 2005;12(4):481-5.
111. Wilczynski NL, McKibbin KA, Walter SD, Garg AX, Haynes RB. MEDLINE clinical queries are robust when searching in recent publishing years. *J Am Med Inform Assoc JAMIA*. abril de 2013;20(2):363-8.
112. Wong SS-L, Wilczynski NL, Haynes RB, Ramkissoonsingh R, Hedges Team. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annu Symp Proc AMIA Symp AMIA Symp*. 2003;728-32.
113. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc JAMIA*. agosto de 2001;8(4):391-7.
114. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. mayo de 2003;56(5):441-7.
115. Peek N, Arts DGT, Bosman RJ, van der Voort PHJ, de Keizer NF. External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol*. mayo de 2007;60(5):491-501.
116. Katz MH. *Multivariable analysis: a practical guide for clinicians*. Cambridge, UK; New York: Cambridge University Press; 1999.
117. Polterauer S, Grimm C, Hofstetter G, Concin N, Natter C, Sturdza A, et al. Nomogram prediction for overall survival of patients diagnosed with cervical cancer. *Br J Cancer*. 4 de septiembre de 2012;107(6):918-24.
118. Tseng J-Y, Yen M-S, Twu N-F, Lai C-R, Horng H-C, Tseng C-C, et al. Prognostic nomogram for overall survival in stage IIB-IVA cervical cancer patients treated with concurrent chemoradiotherapy. *Am J Obstet Gynecol*. febrero de 2010;202(2):174.e1-7.
119. Shim S-H, Lee S-W, Park J-Y, Kim YS, Kim D-Y, Kim J-H, et al. Risk assessment model for overall survival in patients with locally advanced cervical cancer treated with definitive concurrent chemoradiotherapy. *Gynecol Oncol*. enero de 2013;128(1):54-9.

120. Kidd EA, El Naqa I, Siegel BA, Dehdashti F, Grigsby PW. FDG-PET-based prognostic nomograms for locally advanced cervical cancer. *Gynecol Oncol.* octubre de 2012;127(1):136-40.
121. Rose PG, Java J, Whitney CW, Stehman FB, Lanciano R, Thomas GM, et al. Nomograms Predicting Progression-Free Survival, Overall Survival, and Pelvic Recurrence in Locally Advanced Cervical Cancer Developed From an Analysis of Identifiable Prognostic Factors in Patients From NRG Oncology/Gynecologic Oncology Group Randomized Trials of Chemoradiotherapy. *J Clin Oncol Off J Am Soc Clin Oncol.* 1 de julio de 2015;33(19):2136-42.
122. Liang J-A, Chen S-W, Chang W-C, Hung Y-C, Yeh L-S, Yang S-N, et al. Risk stratification for failure in patients with advanced cervical cancer after concurrent chemoradiotherapy: another way to optimise treatment results. *Clin Oncol R Coll Radiol G B.* noviembre de 2008;20(9):683-90.
123. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Disponible en: <http://www.R-project.org>
124. López Encuentra A, Gómez De La Cámara A, Varela De Ugarte A, Mañes N, Llobregat N. [The Will-Rogers phenomenon. Stage migration in bronchogenic carcinoma after applying certainty criteria]. *Arch Bronconeumol.* abril de 2002;38(4):166-71.
125. Steyerberg EW, Van Calster B, Pencina MJ. [Performance measures for prediction models and markers: evaluation of predictions and classifications]. *Rev Esp Cardiol.* septiembre de 2011;64(9):788-94.
126. Harrell FE. rms: Regression Modeling Strategies. R package version 4.4-1. [Internet]. 2015. Disponible en: <https://CRAN.R-project.org/package=rms>
127. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med.* 1 de agosto de 1994; 121(3): 200-6.
128. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-17. 2015.
129. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge: Cambridge University Press; 1997.
130. Steyerberg EW, Vedder MM, Leening MJG, Postmus D, D'Agostino RB, Van Calster B, et al. Graphical assessment of incremental value of novel markers in prediction models: From statistical to decision analytical perspectives. *Biom J Biom Z.* julio de 2015;57(4):556-70.
131. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* abril de 1982;143(1):29-36.

132. Vickers AJ, Cronin AM, Gönen M. A simple decision analytic solution to the comparison of two binary diagnostic tests. *Stat Med.* 20 de mayo de 2013;32(11):1865-76.
133. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford University Press; 2003. 319 p.
134. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 21 de julio de 2009;6(7):e1000100.
135. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 21 de julio de 2009;6(7):e1000097.
136. Seo Y, Yoo SY, Kim M-S, Yang KM, Yoo HJ, Kim JH, et al. Nomogram prediction of overall survival after curative irradiation for uterine cervical cancer. *Int J Radiat Oncol Biol Phys.* 1 de marzo de 2011;79(3):782-7.
137. Yu Q, Lou X-M, He Y. Prediction of local recurrence in cervical cancer by a Cox model comprised of lymph node status, lymph-vascular space invasion, and intratumoral Th17 cell-infiltration. *Med Oncol Northwood Lond Engl.* enero de 2014;31(1):795.
138. Ochi T, Murase K, Fujii T, Kawamura M, Ikezoe J. Survival prediction using artificial neural networks in patients with uterine cervical cancer treated by radiation therapy alone. *Int J Clin Oncol.* octubre de 2002;7(5):294-300.
139. Twiggs LB, Potish RA. Decision theory analysis of the enteric morbidity and surgical staging in the treatment of advanced cervical cancer. *Am J Obstet Gynecol.* 15 de enero de 1984;148(2):134-40.
140. Hong J-H, Tsai C-S, Lai C-H, Chang T-C, Wang C-C, Chou H-H, et al. Risk stratification of patients with advanced squamous cell carcinoma of cervix treated by radiotherapy alone. *Int J Radiat Oncol Biol Phys.* 2005;63(2):492–499.
141. Huang K, Sun H, Li X, Hu T, Yang R, Wang S, et al. Prognostic risk model development and prospective validation among patients with cervical cancer stage IB2 to IIB submitted to neoadjuvant chemotherapy. *Sci Rep.* 9 de junio de 2016;6:27568.
142. Li X, Wei L-C, Zhang Y, Zhao L-N, Li W-W, Ping L-J, et al. The Prognosis and Risk Stratification Based on Pelvic Lymph Node Characteristics in Patients With Locally Advanced Cervical Squamous Cell Carcinoma Treated With Concurrent Chemoradiotherapy: *Int J Gynecol Cancer.* julio de 2016;1.
143. Whitney CW, Sause W, Bundy BN, Malfetano JH, Hannigan EV, Fowler WC, et al. Randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative

- para-aortic lymph nodes: a Gynecologic Oncology Group and Southwest Oncology Group study. *J Clin Oncol Off J Am Soc Clin Oncol*. mayo de 1999;17(5):1339-48.
144. Rose PG, Bundy BN, Watkins EB, Thigpen JT, Deppe G, Maiman MA, et al. Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. *N Engl J Med*. 15 de abril de 1999;340(15):1144-53.
 145. Keys HM, Bundy BN, Stehman FB, Muderspach LI, Chafe WE, Suggs CL, et al. Cisplatin, radiation, and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma. *N Engl J Med*. 15 de abril de 1999;340(15):1154-61.
 146. Lanciano R, Calkins A, Bundy BN, Parham G, Lucci JA, Moore DH, et al. Randomized comparison of weekly cisplatin or protracted venous infusion of fluorouracil in combination with pelvic radiation in advanced cervix cancer: a gynecologic oncology group study. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 de noviembre de 2005;23(33):8289-95.
 147. Thomas G, Ali S, Hoebbers FJP, Darcy KM, Rodgers WH, Patel M, et al. Phase III trial to evaluate the efficacy of maintaining hemoglobin levels above 12.0 g/dL with erythropoietin vs above 10.0 g/dL without erythropoietin in anemic patients receiving concurrent radiation and cisplatin for cervical cancer. *Gynecol Oncol*. febrero de 2008;108(2):317-25.
 148. DiSilvestro PA, Ali S, Craighead PS, Lucci JA, Lee Y-C, Cohn DE, et al. Phase III randomized trial of weekly cisplatin and irradiation versus cisplatin and tirapazamine and irradiation in stages IB2, IIA, IIB, IIIB, and IVA cervical carcinoma limited to the pelvis: a Gynecologic Oncology Group study. *J Clin Oncol Off J Am Soc Clin Oncol*. 10 de febrero de 2014;32(5):458-64.
 149. Gómez de la Cámara A, Tapia PM, Rivas FJP, Rodríguez-Moñino AP. ¿Qué pretenden estimar los estudios de pronóstico? Revisión de la metodología en la investigación pronóstica. *Med Clínica*. 2 de octubre de 2010;135(10):456–461.
 150. Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol*. 15 de octubre de 2010;172(8):971-80.
 151. Moore DH, Tian C, Monk BJ, Long HJ, Omura GA, Bloss JD. Prognostic factors for response to cisplatin-based chemotherapy in advanced cervical carcinoma: a Gynecologic Oncology Group Study. *Gynecol Oncol*. enero de 2010;116(1):44-9.

11 ANEXOS.

11.1 Estrategias de búsqueda.

11.1.1 Estrategia de búsqueda en Medline (pubmed):

Fecha de búsqueda: Octubre de 2016.

Número de referencias encontradas: 30733.

Código de búsqueda: PubMed.

Búsqueda en formato cadena		
#1	(discriminat*[All] AND calibrat*[All])	3964
#2	(external*[All] AND validat*[All])	13119
#3	(nomogram[All Fields] OR nomogram*[All Fields])	6176
#4	(development[All Fields]) AND (index*[All Fields] OR model[All Fields] OR models[All Fields] OR modelling*[All Fields] OR predict[All Fields] OR predicti*[All Fields] OR Prognostic*[All Fields] OR score*[All Fields] OR scoring*[All Fields] OR system[All Fields] OR systems[All Fields] OR refitt*[All] OR Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	1057244
#5	(index*[All Fields]) AND (model[All Fields] OR models[All Fields] OR modelling*[All Fields] OR predict[All Fields] OR predicti*[All Fields] OR Prognostic*[All Fields] OR score*[All Fields] OR scoring*[All Fields] OR system[All Fields] OR systems[All Fields] OR refitt*[All] OR Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	491728
#6	(model[All Fields] OR models[All Fields] OR modelling*[All Fields]) AND (predict[All Fields] OR predicti*[All Fields] OR Prognostic*[All Fields] OR score*[All Fields] OR scoring*[All Fields] OR system[All Fields] OR systems[All Fields] OR refitt*[All] OR Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	1211833
#7	(predict[All Fields] OR predicti*[All Fields] OR Prognostic*[All Fields]) AND (score*[All Fields] OR scoring*[All Fields] OR system[All Fields] OR systems[All Fields] OR refitt*[All] OR Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	443955
#8	(score*[All Fields] OR scoring*[All Fields]) AND (system[All Fields] OR systems[All Fields] OR refitt*[All] OR Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	259238

#9	(system[All Fields] OR systems[All Fields]) AND (refitt*[All] OR Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	374514
#10	(refitt*[All]) AND (Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields] OR rule*[All Fields] OR tool*[All Fields] OR updat*[All])	82
#11	(Risk[All Fields] OR Risks[All Fields] OR Risking*[All Fields]) AND (rule*[All Fields] OR tool*[All Fields] OR updat*[All])	102680
#12	(rule*[All Fields]) AND (tool*[All Fields] OR updat*[All])	7797
#13	(tool*[All Fields]) AND (updat*[All])	6357
#14	Prognosis[MeSH Terms]	1272230
#15	Uterine Cervical Neoplasms[MeSH Terms]	65152
#16	(cervical neoplasms[All Fields] OR cervical neoplas*[All Fields] OR cervical tumour[All Fields] OR cervical tumor[All Fields] OR cervical melanoma[All Fields] OR cervical cancer[All Fields] OR cervical malignan*[All Fields] OR cervical leukemia[All Fields] OR cervical leukaemia[All Fields] OR cervical carcin*[All Fields] OR cervical metastas*[All Fields] OR cervical sarcoma[All Fields])	107311
#17	(cervix neoplasms[All Fields] OR cervix neoplas*[All Fields] OR cervix tumour[All Fields] OR cervix tumor[All Fields] OR cervix melanoma[All Fields] OR cervix cancer[All Fields] OR cervix malignan*[All Fields] OR cervix leukemia[All Fields] OR cervix leukaemia[All Fields] OR cervix carcin*[All Fields] OR cervix metastas*[All Fields] OR cervix sarcoma[All Fields])	76638
#18	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14	3944945
#19	#15 OR #16 OR #17	113943
#20	#18 AND #19	30733

Tabla 16. Estrategia de búsqueda en Medline (pubmed).

11.1.2 Estrategia de búsqueda en ISI Web Of Knowledge:

Fecha de búsqueda: Octubre de 2016.

Número de referencias encontradas: 138770.

Código de búsqueda: ISI Web Of Science.

Búsqueda en formato cadena		
#1	neoplasms OR neoplas* OR tumour OR tumor OR melanoma OR cancer OR malignan* OR leukemia OR leukaemia OR carcin* OR metastas* OR sarcoma	5598137
#2	Uterine Cervical OR uterine cervix OR cervical OR cervix	365158
#3	#2 AND #1	203093
#4	develop* OR development OR index* OR model OR models OR model* OR modelling* OR predict OR predict* OR predicti* OR Prognostic* OR score* OR scoring* OR system OR systems OR refitt* OR Risk* OR Risk OR Risks OR Risking* OR rule* OR tool* OR updat* OR multivar*	32790265
#5	(discriminat* AND calibrat*) OR (external* AND validat*) OR nomogram OR nomogram* OR (prognostic model)	108486
#6	#5 OR #4	32796251
#7	#6 AND #3	138770

Tabla 17. Estrategia de búsqueda en ISI Web Of Knowledge.

11.1.3 Estrategia de búsqueda en EMBASE (Ovid):

Fecha de búsqueda: Octubre de 2016.

Número de referencias encontradas: 9439.

Código de búsqueda: EMBASE.

Búsqueda en formato cadena		
#1	'uterine cervix cancer'/exp	75823
#2	(cervix NEAR/4 (neoplasms OR neoplas* OR tumour OR tumor OR melanoma OR cancer OR malignan* OR leukemia OR leukaemia OR carcin* OR metastas* OR sarcoma)):ab,ti	22231
#3	(cervical NEAR/4 (neoplasms OR neoplas* OR tumour OR tumor OR melanoma OR cancer OR malignan* OR leukemia OR leukaemia OR carcin* OR metastas* OR sarcoma)):ab,ti	72600
#4	('incidence'/de OR 'mortality'/de OR 'follow up'/de OR 'prognosis'/de OR 'prediction and forecasting'/exp OR 'cohort analysis'/exp OR incidence:ab,ti OR mortality:ab,ti OR 'follow up':ab,ti OR prognos*:ab,ti OR predict*:ab,ti OR cohort:ab,ti)	4836492
#5	'prognosis'/exp	549998
#6	(discriminat*:ab AND calibrat*:ab)	5030
#7	(external*:ab AND validat*:ab)	157177
#8	(nomogram:ab,ti OR nomogram*:ab,ti)	8327
#9	(development NEAR/4 (index* OR model* OR predict* OR prognostic* OR score* OR scoring* OR system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	157149
#10	(index* NEAR/4 (model* OR predict* OR prognostic* OR score* OR scoring* OR system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	72306
#11	(model* NEAR/4 (predict* OR prognostic* OR score* OR scoring* OR system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	264343
#12	(predict* NEAR/4 (prognostic* OR score* OR scoring* OR system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	139010
#13	(prognostic* NEAR/4 (score* OR scoring* OR system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	22570
#14	(scoring* NEAR/4 (system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	37743
#15	(score* NEAR/4 (system OR systems OR refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	45884
#16	(system NEAR/4 (refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	20122
#17	(systems NEAR/4 (refitt* OR risk* OR rule* OR tool* OR updat*)):ab,ti	7089
#18	(refitt* NEAR/4 (risk* OR rule* OR tool* OR updat*)):ab,ti	16
#19	(risk* NEAR/4 (rule* OR tool* OR updat*)):ab,ti	15522

#20	(rule* NEAR/4 (tool* OR updat*)):ab,ti	995
#21	(tool* NEAR/4 updat*):ab,ti	422
#22	#1 OR #2 OR #3	110109
#23	#5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21	1194700
#24	#22 AND #4 AND #23	9439

Tabla 18. Estrategia de búsqueda en EMBASE (Ovid).

11.1.4 Estrategia de búsqueda en IME (Índice Médico Español):

Fecha de búsqueda: Junio de 2015.

Número de referencias encontradas: -.

Código de búsqueda: Búsqueda simple IME.

Búsqueda en formato cadena		
#1	nomograma	7
#2	nomogramas	27
#3	nomograma AND cáncer AND cervix	0
#4	pronóstico AND cáncer AND cervix	18
#5	modelo AND cáncer AND cervix	2
#6	predicción AND cáncer AND cervix	0
#7	puntuación AND cáncer AND cervix	0
#8	nomograma AND cáncer AND cervical	0
#9	pronóstico AND cáncer AND cervical	17
#10	modelo AND cáncer AND cervical	5
#11	predicción AND cáncer AND cervical	1
#12	puntuación AND cáncer AND cervical	1
#13	modelos AND pronóstico	51
#14	modelos AND pronóstico	78
#15	validación AND cervix	1
#16	validacion AND cervical	4
#17	validación AND cáncer	41
#18	desarrollo AND cervix	11
#19	cáncer cévico-uterino	10

Tabla 19. Estrategia de búsqueda en IME (Índice Médico Español).

11.1.5 Estrategia de búsqueda en Epistemonikos:

Fecha de búsqueda: Junio de 2015.

Número de referencias encontradas: -.

Código de búsqueda: -.

Búsqueda en formato cadena		
#1	cancer cervix	347
#2	nomogram	127
#3	nomograms	39
#4	cervical cancer	1894
#5	prognostic model	1146
#6	score model	4030
#7	external validation	135

Tabla 20. Estrategia de búsqueda en Epistemonikos.

11.1.6 Estrategia de búsqueda en IBECS (isciii):

Fecha de búsqueda: Junio de 2015.

Número de referencias encontradas: -.

Código de búsqueda: -.

Búsqueda en formato cadena		
#1	cervical AND cancer AND prognos\$	54
#2	validacion AND externa	37
#3	cervix AND cancer AND prognos\$	36
#4	nomogra\$	40
#5	score model\$ cervix	1
#6	model\$ cancer cervi\$	34

Tabla 21. Estrategia de búsqueda en IBECS (isciii).

11.1.7 Estrategia de búsqueda en LILACS:

Fecha de búsqueda: Junio de 2015.

Número de referencias encontradas: -.

Código de búsqueda: .

Búsqueda en formato cadena		
#1	tw:(tw:(cancer cervix model) AND (instance: "regional")) AND (instance: "regional") AND (db:("LILACS" OR "coleccionaSUS" OR "IBECs" OR "BDENF" OR "HISA" OR "MedCarib" OR "PAHO"))	31
#2	tw:(cancer cervix nomogram)	3

Tabla 22. Estrategia de búsqueda en LILACS.

11.2 Informe del CEIC del Hospital Universitario 12 de Octubre.

Nº CEIC: 15/210

INFORME DEL COMITÉ ÉTICO DE INVESTIGACION CLINICA

Dña. MARIA UGALDE DIEZ, Secretaria del Comité Ético de Investigación Clínica (CEIC) del Hospital Universitario Doce de Octubre.

CERTIFICA:

Que este Comité, en la reunión celebrada el día **30/06/2015**, ha evaluado los aspectos éticos del Proyecto de Investigación titulado:

**MODELOS MATEMATICOS DE PRONOSTICO PARA CANCER DE CERVIX LOCALMENTE AVANZADO.
VALIDACION EXTERNA DE LOS MODELOS EXISTENTES**

Del cual el Sr. **David LORA PABLOS**,

Del Servicio de **UNIDAD DE INVESTIGACION** es el Investigador Principal

Entendiendo que dicho estudio se ajusta a las normas éticas esenciales y criterios deontológicos que rigen en este Centro, cumpliendo los requisitos metodológicos necesarios, este Comité **INFORMA FAVORABLEMENTE** a la realización de dicho proyecto en este Centro.

Lo que firmo en Madrid, a 2 de julio de 2015,



Hospital Universitario
12 de Octubre
Comunidad de Madrid
COMITÉ DE ÉTICA DE LA INVESTIGACIÓN
CON MEDICAMENTOS - CEIM

Firmado: *Dra. Maria Ugalde Diez*
Secretaria CEIC Hospital 12 de Octubre.

Nº CEIC: 15/210

TÍTULO: *MODELOS MATEMATICOS DE PRONOSTICO PARA CANCER DE CERVIX LOCALMENTE AVANZADO. VALIDACION EXTERNA DE LOS MODELOS EXISTENTES*

Investigador Principal: LORA PABLOS, David

D^a MARIA UGALDE DIEZ, SECRETARIA DEL COMITÉ ETICO DE INVESTIGACION CLINICA DEL HOSPITAL UNIVERSITARIO 12 DE OCTUBRE DE MADRID

Hace constar que:

1. En la reunión ordinaria celebrada el día **30/06/2015**, se decidió emitir el informe correspondiente al proyecto de Investigación de referencia.
2. El CEIC del Hospital Universitario 12 de Octubre, tanto en su composición como en sus PNTs, cumple con las normas de BPC (CPMP/ICH/135/95)
3. La composición del Ceic del Hospital Universitario 12 de Octubre que evaluó el proyecto fue la siguiente:

PRESIDENTE	Dra. M ^a del Puy Goyache Goñi	Farmacéutico Adjunto de Farmacia Hospitalaria
VICEPRESIDENTA	Dra. Carmen Jimenez López-Guarch	Medico Cardiología
SECRETARIA	Dra. María Ugalde Díez	Dra. en Ciencias Biológicas
VOCALES:	D ^a M ^a Luisa Albelda de la Haza	Licenciada Derecho
	Dra. M ^a de las Mercedes Catalán Gómez	Medico Medicina Intensiva
	Dra. Eva M ^a Ciruelos Gil	Medico Oncología
	Dr. Rafael Delgado Vazquez	Jefe de Sección de Microbiología Molecular
	Dr. Alberto Galindo Izquierdo	Jefe de Sección Obstetricia y Ginecología
	Dr. Eduardo Gutiérrez Martínez	Medico Nefrología
	Dra. María del Pilar Martínez Sanchez	Medico Hematología
	Sr. Francisco Javier Mazuecos Gómez	Auxiliar Administrativo Endoscopia
	Dr. Cesar Minué Lorenzo	Médico de Familia de Atención Primaria
	Dr. José Manuel Moreno Villares	Medico Pediatría
	Dra. Gloria Orejón de Luna	Pediatra de Atención Primaria
	D ^a . Montserrat Pilas Pérez	Diplomado Universitario en Enfermería
	Dr. Santiago Ponce Aix	Medico Oncología
	Sra. Yolanda del Rey Granada	Diplomado Universitario en Enfermería
	Dra. María del Carmen Riesco Martínez	Medico Oncología
	Dra. Yolanda Rodríguez Gil	Médico Anatomía Patológica
	Dr. Roberto Rodríguez Jimenez	Médico Psiquiatría
	Dra. Belen Ruiz Antoran	Farmacóloga clínica (Clínica Puerta de Hierro)
	Dra. Raquel Siguín Gómez	Farmacéutica Atención Primaria
	D ^a Rosa M ^a Vega Viaña	Plataforma de Ensayos Clínicos SCReN
	Dr. Alberto Villarejo Galende	Medico Neurología

Que en el caso de que se evaluara algún proyecto del que un miembro sea investigador/colaborador, este se ausentara de la reunión durante la discusión del proyecto.

Para que conste donde proceda, y a petición del promotor,

11.3 Características de los estudios incluidos en la revisión.

X Li (2016)(142)	
Tipo de estudio	Estudio de cohorte, recogida de la información retrospectiva.
Año	2011-2014
Localización	China (1 centros)
Participantes	Estadio FIGO: II; III; IVA
Tratamiento	Quimiorradioterapia radical (cisplatino o cisplatino más paclitaxel).
Criterios de inclusión	Pacientes con carcinoma de células escamosas. Valores de ECOG de 0 o 1. Estadio FIGO de II, III o IVA. Sin metástasis de los ganglios linfáticos de la paraaórtica. Sin operación, tratamiento neoadyuvante o adyuvante de quimioterapia.
Muestra	300 pacientes
Desenlace	Muerte causa específica y recurrencia a distancia.

Tabla 23. Características del estudio X Li (2016)(142).

SH Shim (2013) (119)	
Tipo de estudio	Estudio de cohorte, recogida de la información retrospectiva.
Año	1998-2008
Localización	Corea del Sur (Un único centro)
Participantes	Estadio FIGO: IB2; IIA; IIB; IIIA; IIIB. 4221 pacientes.
Tratamiento	Quimioterapia basada en platino con radioterapia de campo extendido. Tres tipos: 1) 5-Fluorouracil + Cisplatino; 2) Paclitaxel + Cisplatino; 3) Paclitaxel + carboplatino.
Criterios de inclusión	<p>Estadificación clínica con resultados quirúrgicos. Edad entre 18 y 80 años. Estadio FIGO: IB2-IVA.</p> <p>Tratamiento de quimioradioterapia basado en platino sin quimioterapia de consolidación.</p> <p>En caso de existir uso de Fluorodesosiglucosa en la tomografía por emisión de positrones que sea anterior a la quimioradioterapia. Sin evidencia de metástasis a distancia. ECOG: 0, 1 o 2.</p> <p>Sin tratamiento previo de quimioterapia o radioterapia. Sin enfermedad que afecte a la supervivencia.</p>
Muestra	209 pacientes incluido en la muestra de desarrollo.
Desenlace	Muerte por causa específica.

Tabla 24. Características del estudio SH Shim (2013) (119).

S Kang (2012)(41)	
Tipo de estudio	Estudio de cohorte, recogida de la información retrospectiva.
Año	2001-2009
Localización	Corea del Sur (3 centros para desarrollo; 2 para validación).
Participantes	Estadio FIGO: IB; IIA; IIB; III; IVA. 748 Pacientes candidatos para desarrollo. 167 pacientes candidatos para validación.
Tratamiento	Quimioterapia con radioterapia basada en platino con intención de curar.
Criterios de inclusión	Diagnóstico histológico de cáncer de cérvix primario. Estadio FIGO: IIB-IVA o IB2-IIA bulky. No se incluyen IB1 o IVB. Mujeres tratadas con quimioradioterapia basada en platino de intención curativa con o sin radioterapia de campo extendido. En caso de existir uso de FDG/CT o PET/CT anterior a la quimioradioterapia, que no sea más de 4 semanas. Sin tratamiento previo de quimioterapia o radioterapia. Sin otro cáncer concomitante. Sin tipo histológico de carcinoma de células pequeñas, de tipo neuroendocrino o otros tipos histológicos raros. Dosis de radioterapia mayor a 40Gy.
Muestra	434 pacientes incluido en la muestra de desarrollo. 115 pacientes incluido en la muestra de validación.
Desenlace	Recurrencia a distancia.

Tabla 25. Características del estudio S Kang (2012)(41).

JY Tseng (2010) (118)

Tipo de estudio	Estudio de cohorte prospectivo
Año	1999-2006
Localización	Taiwan (Un único centro)
Participantes	Estadio FIGO: IIB; IIIA; IIIB; IVA. 284 pacientes
Tratamiento	Quimioterapia basada en cisplatino o carboplatino más radioterapia externa en la zona de la pelvis seguida de braquiterapia intracavitaria. En algunos casos, se hizo quimioterapia de consolidación, tres posibilidades: Ifofamide más cisplatino; Bleomycin más ifosfamide más cisplatino; Paclitaxel más cisplatino.
Criterios de inclusión	Confirmación histológica por biopsia o conización con asa electroquirúrgica. Estadio FIGO: IIB-IVA. Pueden ser tratados con quimioterapia de consolidación. Pueden tener sospecha de metástasis en ganglios paraaórticos. ECOG: 0, 1 o 2. Pacientes con carcinoma de células escamosas.
Muestra	251 pacientes incluido en la muestra de desarrollo.
Desenlace	Muerte por causa específica.

Tabla 26. Características del estudio JY Tseng (2010) (118).

JA Liang (2008)(122)

Tipo de estudio	Estudio de cohorte, recogida de la información retrospectiva.
Año	2001-2006
Localización	Taiwan (Un único centro)
Participantes	Estadio FIGO: IB2; IIA; IIB; III; IVA. 163 Pacientes.
Tratamiento	Radioterapia externa seguida de braquiterapia intracavitaria más quimioterapia basada en cisplatino.
Criterios de inclusión	Estadio FIGO: IB2-IVA. Mujeres tratadas con quimioterapia de intención curativa. Sin evidencia de metástasis en los ganglios linfáticos paraaórticos.
Muestra	148 pacientes incluido en la muestra de desarrollo.
Desenlace	Metástasis (Libre de enfermedad). Recurrencia para-aórtica.

Tabla 27. Características del estudio JA Liang (2008)(122).

S Polterauer (2012)(117)	
Tipo de estudio	Estudio de cohorte prospectiva.
Año	1996-2009
Localización	Austria (Dos centros).
Participantes	Estadio FIGO: IA1; IA2; IB1; IB2; IIA; IIB; IIIA; IIIB; IVA. 692 Paciente.
Tratamiento	Según estadio FIGO. IA1: conización o histerectomía simple. A las pacientes con invasión del espacio vascular linfático, se les hace linfadenectomía pélvica y/o paraaórtica. IB1-IIA: histerectomía radical o trachelectomía más linfadenectomía pélvica y/o paraaórtica. Aquellas con ganglios linfáticos positivos, invasión del parametrio y márgenes quirúrgicos positivos son tratadas con radioterapia adjuvante postoperatoria con o sin quimioterapia concurrente. IIB-IV: Quimioradioterapia. Se hace linfadenectomía pélvica y/o paraaórtica antes de la quimioradioterapia. Radiación del campo extendido en pacientes con ganglios positivos en la región iliaca.
Criterios de inclusión	Todas las mujeres registradas.
Muestra	528 pacientes incluido en la muestra de desarrollo.
Desenlace	Muerte por causa específica.

Tabla 28. Características del estudio S Polterauer (2012)(117).

EA Kidd (2012)(120)	
Tipo de estudio	Estudio de cohorte, recogida de la información retrospectiva.
Año	1998-2008
Localización	EEUU (Un único centro).
Participantes	Estadio FIGO: IB1; IB2; IIA; IIB; IIIA; IIIB; IVA. 482 pacientes.
Tratamiento	Dos grupos: Tratamiento 1: Pacientes tratadas con radioterapia definitiva, incluyendo irradiación en pelvis y braquiterapia intracavitaria. Tratamiento 1 más quimioterapia concurrente con cisplatino. Todas las pacientes reciben FDG-PET o PET/TC antes del tratamiento.
Criterios de inclusión	Diagnóstico primario de cáncer de cérvix. Radioterapia definitiva o quimioradioterapia. Todas las pacientes reciben FDG-PET o PET/TC antes del tratamiento.
Muestra	234 pacientes incluido en la muestra de desarrollo.
Desenlace	Muerte. Muerte por causa específica. Recurrencia.

Tabla 29. Características del estudio EA Kidd (2012)(120).

PG Rose (2015)(121)

Tipo de estudio	Ensayo clínico. Estudio pronóstico retrospectivo.
Año	1986-2009*
Localización	Canada - EEUU (Seis ensayos clínicos).
Participantes	Estadio FIGO: IB; IIA; IIB; IIIA; IIIB; IVA. 2041 Pacientes.
Tratamiento	<p>CW Whitney (1999)(143). Dos ramas de tratamientos: Tratamiento 1: Radioterapia externa pélvica con infusión de fluorouracilo concurrente y bolos de cisplatino. Tratamiento 2: Radioterapia externa pélvica con oral hifoxiurea. A todas las pacientes se les hace linfadenectomía paraaórtica.</p> <p>PG Rose (1999)(144). A todos los pacientes se les hace linfadenectomía peritoneal paraaórtica. Tres ramas de tratamientos: Tratamiento 1: Radioterapia aplicada a la región pélvica más cisplatino. Tratamiento 2: Radioterapia aplicada a la región pélvica más cisplatino, más fluorouracilo y más hidroxurea. Tratamiento 3: Radioterapia aplicada a la región pélvica más hidroxurea.</p> <p>HM Keys (1999)(145). Se les hace histerectomía auxiliar. Dos ramas de tratamientos: Tratamiento 1: Radioterapia sola (incluyendo irradiación externa, braquiterapia intracavitaria e histerectomía) en pelvis y braquiterapia intracavitaria). Tratamiento 2: Radioterapia en combinación de cisplatino.</p> <p>R Lanciano (2005)(146). Pacientes sin cirugía de los ganglios paraaórticos. Dos ramas de tratamientos: Tratamiento 1: Radioterapia más infusión venosa prolongada de fluororacil. Tratamiento 2: Radioterapia más cisplatino.</p> <p>G Thomas (2008) (147). Todas las mujeres recibieron radioterapia más quimioterapia basada en cisplatino. Dos ramas de tratamientos: Tratamiento 1: Eritropoyetina recombinante humana. Tratamiento 2: Transfusión estándar por debajo de unos niveles.</p> <p>PA DiSilvestro (2014) (148). Dos ramas de tratamientos: Tratamiento 1: Quimioterapia con cisplatino. Tratamiento 2: Quimioterapia con cisplatino más tirapazanine.</p>

Criterios de inclusión	<p>CW Whitney (1999)(143). Biopsia a todos los participantes. Estadio FIGO: IIB; III o IVA. Función normal de riñón, hígado y médula ósea. Sin irradiación de pelvis o quimioterapia citotóxica. Libres de infecciones. Sin contraindicaciones para operar. ECOG inferior o igual a 3. Sin otros cánceres concomitantes. Sin metástasis en ganglios paraaórticos o intraabdominales o pacientes con lavados citológicos positivos. Tipo histológico de carcinoma de células escamosas, adenocarcinoma o adenoescamoso. PG Rose (1999)(144). Confirmación histológica. Estadio FIGO: IIB; III o IVA. ECOG inferior o igual a 3. Sin otros cánceres concomitantes. Sin enfermedad fuera de la pelvis, metástasis en ganglios linfáticos, paraaórticos o enfermedad retroperitoneal. Tipo histológico de carcinoma de células escamosas, adenocarcinoma o adenoescamoso. HM Keys (1999)(145). Estadio FIGO: IB. Función normal de riñón, hígado y médula ósea. Sin tratamiento previo de radioterapia o quimioterapia. Sin contraindicaciones para operar. ECOG inferior o igual a 3. Sin otros cánceres concomitantes. Sin afectación en ganglios linfáticos. Tipo histológico de carcinoma de células escamosas, adenocarcinoma o adenoescamoso. R Lanciano (2005)(146). Estadio FIGO: IIB; IIIB o IVA. (No se incluyen IIIA). Función normal de riñón, hígado y médula ósea. Sin tratamiento previo. Tipo histológico de carcinoma de células escamosas, adenocarcinoma o adenoescamoso. G Thomas (2008) (147). Estadio FIGO: IIB; IIIB o IVA. Función normal de riñón, hígado y médula ósea. Sin ganglios para-aórticos positivos. PA DiSilvestro (2014) (148). Estadio FIGO: IB2 o IIA (bulky >4cm); o IIB; IIIB o IVA. Función normal de riñón, hígado y médula ósea. ECOG inferior o igual a 3. Sin otros cánceres concomitantes. Sin ganglios linfáticos paraaórticos o metástasis a distancia. Tipo histológico de carcinoma de células escamosas, adenocarcinoma o adenoescamoso.</p>
Muestra	2041 pacientes incluido en la muestra de desarrollo.

Desenlace

**Muerte.
Recurrencia.**

Tabla 30. Características del estudio PG Rose (2015)(121).

11.4 Guía TRIPOD.

La reproducción de la guía TRIPOD (47,89) está regido por los derechos contenidos en:

47. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Diabet Med J Br Diabet Assoc.* febrero de 2015;32(2):146-54.
89. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 6 de enero de 2015;162(1):W1-73.

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	
	5b	Describe eligibility criteria for participants.	
	5c	Give details of treatments received, if relevant.	
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	Explain how the study size was arrived at.	
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	
Risk groups	11	Provide details on how risk groups were created, if done.	
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	
Model development	14a	Specify the number of participants and outcome events in each analysis.	
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	
	15b	Explain how to use the prediction model.	
Model performance	16	Report performance measures (with CIs) for the prediction model.	
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	
Implications	20	Discuss the potential clinical use of the model and implications for future research.	
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	
Funding	22	Give the source of funding and the role of the funders for the present study.	

12 PUBLICACIÓN

Notwithstanding the Stata Journal copyright notice, StataCorp and the Stata Journal hereby agree that David Lora will be entitled to include the article "Features of the area under the receiver operating characteristic (ROC) curve. A good practice." (Lora, Contador, Perez-Regaders, and Gomez de la Camara, Stata Journal 16: 185-196) in his thesis. A total of 25 copies of this article may be distributed.

THE STATA JOURNAL

Editors

H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy
MAARTEN L. BUIS, University of Konstanz, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for
Medical Sciences
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
DAVID EPSTEIN, Columbia University
ALLAN GREGORY, Queen's University
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and
Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
JENS LAURITSEN, Odense University Hospital
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSON, Imperial College, London
AUSTIN NICHOLS, Urban Institute, Washington DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC Clinical Trials Unit,
London
PHILIP RYAN, University of Adelaide
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh
JEROEN WEESIE, Utrecht University
IAN WHITE, MRC Biostatistics Unit, Cambridge
NICHOLAS J. G. WINTER, University of Virginia
JEFFREY WOOLDRIDGE, Michigan State University

Stata Press Editorial Manager

LISA GILMORE

Stata Press Copy Editors

DAVID CULWELL, SHELBI SEINER, and DEIRDRE SKAGGS

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		Printed & electronic	
1-year subscription	\$115	1-year subscription	\$145
2-year subscription	\$210	2-year subscription	\$270
3-year subscription	\$285	3-year subscription	\$375
1-year student subscription	\$ 85	1-year student subscription	\$115
1-year institutional subscription	\$345	1-year institutional subscription	\$375
2-year institutional subscription	\$625	2-year institutional subscription	\$685
3-year institutional subscription	\$875	3-year institutional subscription	\$965
Electronic only		Electronic only	
1-year subscription	\$ 85	1-year subscription	\$ 85
2-year subscription	\$155	2-year subscription	\$155
3-year subscription	\$215	3-year subscription	\$215
1-year student subscription	\$ 55	1-year student subscription	\$ 55

Back issues of the *Stata Journal* may be ordered online at

<http://www.stata.com/bookstore/sjj.html>

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

<http://www.stata-journal.com/archives.html>

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.



Copyright © 2016 by StataCorp LP

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, **STATA**, Stata Press, Mata, **MATA**, and NetCourse are registered trademarks of StataCorp LP.

Features of the area under the receiver operating characteristic (ROC) curve. A good practice.

David Lora
Clinical Research Unit (imas12)
Hospital Universitario 12 de Octubre
and CIBER de Epidemiología y Salud Pública (CIBERESP)
Madrid, Spain
david@h12o.es

Israel Contador
Department of Basic Psychology, Psychobiology and Methodology of
Behavioral Sciences
University of Salamanca
Salamanca, Spain

José F. Pérez-Regadera
Department of Radiation Oncology
Hospital Universitario 12 de Octubre
Madrid, Spain

Agustín Gómez de la Cámara
Clinical Research Unit (imas12)
Hospital Universitario 12 de Octubre
and CIBER de Epidemiología y Salud Pública (CIBERESP)
Madrid, Spain

Abstract. The area under the receiver operating characteristic (ROC) curve is a measure of discrimination ability used in diagnostic and prognostic research. The ROC plot is usually represented without additional information about decision thresholds used to generate the graph. In our article, we show that adding at least one or more informative cutoff points on the ROC graph facilitates the characterization of the test and the evaluation of the discriminatory capacities, which can result in more informed medical decisions. We use the `rocreg` and `rocregplot` commands.

Keywords: st0430, receiver operating characteristic (ROC) curve, area under the ROC curve, cervix cancer, diagnostic test, discrimination, prognostic models, `rocreg`, `rocregplot`

1 Introduction

The receiver operating characteristic (ROC) area represents the probability that in a specific diagnostic test or prognostic model, a randomly chosen diseased subject is ranked with greater suspicion than a randomly chosen nondiseased subject (Hanley and McNeil 1982). The area under the ROC curve is a measure of discrimination ability used in diagnostic tests and prognostic models. The discriminatory capacity corresponds to the area below the curve. The ROC area can be graphed by varying the cutoff points used to determine which values of the clinical procedure will be considered abnormal and then plotting the resulting true-positive rate (sensitivity) against the corresponding false-positive rate ($1 - \text{specificity}$) (DeLong, DeLong, and Clarke-Pearson 1988). However, decision thresholds are usually not displayed on the ROC plot, though they are known and used to generate the graph (Zweig and Campbell 1993). In this article, we show that the addition of cutoff point information enables one to 1) locate the decision threshold with respect to its sensitivity and specificity and 2) show its importance regarding the other thresholds. This information, which is useful in making medical decisions (Royston, Altman, and Sauerbrei 2006), facilitates the characterization of the test as well as the evaluation of the discriminatory capacity when using the ROC plot (Moons et al. 2015).

2 Example

To demonstrate the importance of the description of the area under the ROC curve, we generated a random dataset of 200 patients in the early stages of cervical cancer with three variables: the lymph node metastasis (LNM) status (40 out of 200) and the predicted probabilities of two preoperative prognostic models for the identification of LNM in early cervical cancer. The first model incorporated age, tumor size by magnetic resonance imaging (MRI), and LNM assessed by positron emission tomography and computed tomography variables; the second model incorporated age, tumor size by MRI, and LNM assessed by MRI and squamous cell carcinoma antigen. The sample was formed by comparing the patients' data with the information given in Kim et al. (2014).

```
. set obs 200
number of observations (_N) was 0, now 200
. set seed 1556
. generate LNM=1 if _n<=40
(160 missing values generated)
. generate predprob=0.01*runiform() if _n<=1
(199 missing values generated)
. replace predprob=0.01+(0.05-0.01)*runiform() if _n<=3 & _n>1
(2 real changes made)
. replace predprob=0.05+(0.25-0.05)*runiform() if _n<=10 & _n>3
(7 real changes made)
. replace predprob=0.25+(0.50-0.25)*runiform() if _n<=18 & _n>10
(8 real changes made)
. replace predprob=0.50+(0.75-0.50)*runiform() if _n<=28 & _n>18
(10 real changes made)
```

```

. replace predprob=0.75+(1.00-0.75)*runiform() if _n<=40 & _n>28
(12 real changes made)
. generate predprob2=0.15*runiform() if _n<=1
(199 missing values generated)
. replace predprob2=0.15+(0.25-0.15)*runiform() if _n<=3 & _n>1
(2 real changes made)
. replace predprob2=0.25+(0.45-0.25)*runiform() if _n<=10 & _n>3
(7 real changes made)
. replace predprob2=0.45+(0.65-0.45)*runiform() if _n<=18 & _n>10
(8 real changes made)
. replace predprob2=0.65+(0.85-0.65)*runiform() if _n<=28 & _n>18
(10 real changes made)
. replace predprob2=0.85+(1.00-0.85)*runiform() if _n<=40 & _n>28
(12 real changes made)
. replace LNM=0 if _n> 40
(160 real changes made)
. replace predprob=0.01*runiform() if _n<=86 & _n>40
(46 real changes made)
. replace predprob=0.01+(0.05-0.01)*runiform() if _n<=124 & _n>86
(38 real changes made)
. replace predprob=0.05+(0.25-0.05)*runiform() if _n<=161 & _n>124
(37 real changes made)
. replace predprob=0.25+(0.50-0.25)*runiform() if _n<=180 & _n>161
(19 real changes made)
. replace predprob=0.50+(0.75-0.50)*runiform() if _n<=195 & _n>180
(15 real changes made)
. replace predprob=0.75+(1.00-0.75)*runiform() if _n<=200 & _n>195
(5 real changes made)
. replace predprob2=0.15*runiform() if _n<=86 & _n>40
(46 real changes made)
. replace predprob2=0.15+(0.25-0.15)*runiform() if _n<=124 & _n>86
(38 real changes made)
. replace predprob2=0.25+(0.45-0.25)*runiform() if _n<=161 & _n>124
(37 real changes made)
. replace predprob2=0.45+(0.65-0.45)*runiform() if _n<=180 & _n>161
(19 real changes made)
. replace predprob2=0.65+(0.85-0.65)*runiform() if _n<=195 & _n>180
(15 real changes made)
. replace predprob2=0.85+(1.00-0.85)*runiform() if _n<=200 & _n>195
(5 real changes made)

```

Using the `rocreg` command, we estimate the area under the ROC curve for the classifier based on model 1 with 95% confidence intervals (CI) and store the true-positive rate (sensitivity) and false-positive rate ($1 - \text{specificity}$) values in variables for each classifier point, `_roc_predprob` and `_fpr_predprob`, respectively.

```
. rocreg LNM predprob, nodots
Bootstrap results                               Number of obs   =       200
                                                Replications   =     1,000

Nonparametric ROC estimation
Control standardization: empirical
ROC method           : empirical

Area under the ROC curve
Status      : LNM
Classifier: predprob
```

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
	.8335938	.0012301	.0341697	.7666225	.900565 (N)
				.7586384	.8939733 (P)
				.7467949	.8849932 (BC)

We use the `rocregplot` command to draw the ROC curve for the first model (figure 1).

```
. rocregplot, plotlopts(msymbol(none) lcolor(black))
> scheme(sicolor) rlopts(lcolor(black))
> ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
> xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
> legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))
```

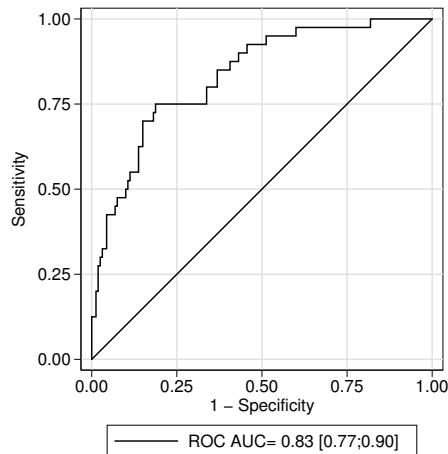


Figure 1. ROC curve for model 1

This curve is depicted by consecutive values from the test (that is, predicted probabilities from the prognostic model) to classify an individual as positive (LNM ill = value is above the threshold) or negative (free of LNM = values are low or under the selected threshold) in comparison with a gold-standard criterion. An increase of the cutoff point

would decrease sensitivity and increase specificity. We have added the model's discriminatory capacity—0.83 (95% CI [0.77; 0.90])—to the plot using the `legend()` option. We can also characterize and enhance the model by displaying cutoff point information on the plot. To add the predicted probabilities, we first list the observations with predicted probabilities close to 1%, 5%, and 10%. The observations with predicted probabilities closest to 1%, 5%, and 10% are indicated with an arrow.

```
. sort predprob
. list if (predprob > 0.009 & predprob < 0.013)
```

	LNM	predprob	predpr-2	_roc_p-b	_fpr_p-b	
45.	0	.0095514	.0902184	.975	.73125	
46.	0	.0095826	.0040212	.975	.725	
47.	0	.0096395	.1032082	.975	.71875	<--
48.	0	.0120601	.2063762	.975	.7125	
49.	0	.0128457	.184004	.975	.70625	

```
. list if (predprob > 0.049 & predprob < 0.06)
```

	LNM	predprob	predpr-2	_roc_p-b	_fpr_p-b	
85.	0	.0490097	.2001811	.925	.49375	
86.	0	.0498729	.228905	.925	.4875	
87.	0	.0499744	.2237535	.925	.48125	<--
88.	0	.0567622	.3069373	.925	.475	

```
. list if (predprob > 0.09 & predprob < 0.11)
```

	LNM	predprob	predpr-2	_roc_p-b	_fpr_p-b	
102.	0	.0936247	.3390107	.875	.40625	<--
103.	0	.1064184	.2700449	.85	.4	

Now, we use `rocregplot` to redraw the ROC curve. We add captions at the sensitivity and $1 -$ specificity of the indicated observations to the plot using the `text()` option (figure 2).

```
. rocregplot, plotlopts(msymbol(none) lcolor(black))
>   scheme(sicolor) rlopts(lcolor(black))
>   ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
>   xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
>   text(0.99 0.69 "1%", size(small) placement(n))
>   text(0.94 0.44 "5%", size(small) placement(n))
>   text(0.89 0.37 "10%", size(small) placement(n))
>   text(0.98 0.72 "x", size(large) placement(c))
>   text(0.93 0.48 "x", size(large) placement(c))
>   text(0.88 0.41 "x", size(large) placement(c))
>   legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))
```

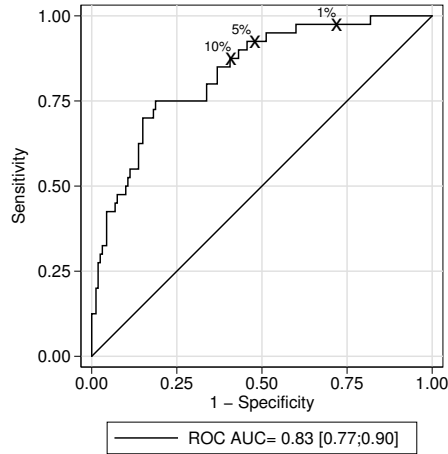


Figure 2. ROC curve for model 1 with predicted probabilities closest to 1%, 5%, and 10%

The result is that we show the predicted probabilities of developing LNM with the associated sensitivity and specificity values in the ROC plot. Further, we can describe the sensitivity and 95% CI with respect to its 1 – specificity using the `roc()` option (figure 3) and describe the false-positive rate and 95% CI for given true-positive values using the `invroc()` option in the `rocreg` command (figure 4):

```
. rocreg LNM predprob, nodots roc(0.71875 0.48125 0.40625)
Bootstrap results                               Number of obs   =       200
                                                Replications    =       1,000

Nonparametric ROC estimation
Control standardization: empirical
ROC method           : empirical
ROC curve
  Status      : LNM
  Classifier: predprob
```

ROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
.71875	.975	-.0002608	.0249877	.9260249	1.023975 (N)
				.9166667	1 (P)
				.9090909	1 (BC)
.48125	.925	-.0044098	.0487009	.829548	1.020452 (N)
				.8101673	1 (P)
				.8	1 (BC)
.40625	.875	-.0159269	.0697107	.7383696	1.01163 (N)
				.7105263	.9725976 (P)
				.7333333	.9767442 (BC)

```

. rocregplot, btype(p) plotlopts(msymbol(none) lcolor(black))
> scheme(s1color) rlopts(lcolor(black))
> ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
> xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
> legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))

```

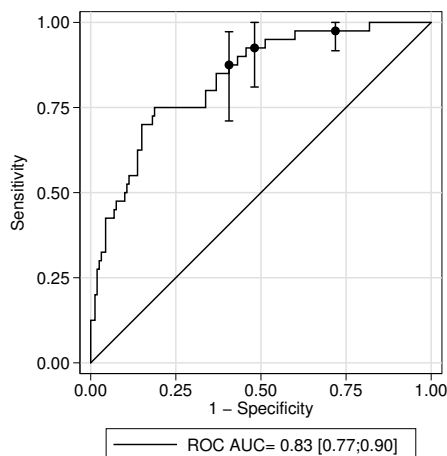


Figure 3. ROC curve for model 1 showing sensitivity and 95% CI with respect to its 1 – specificity

```

. rocreg LNM predprob, nodots invroc(0.875 0.925 0.975)
Bootstrap results                Number of obs    =      200
                                Replications        =      1,000

Nonparametric ROC estimation
Control standardization: empirical
ROC method                      : empirical

False-positive rate
Status       : LNM
Classifier   : predprob

```

invROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
.875	.40625	.0025914	.0776321	.254094	.558406 (N)
				.1911568	.5642543 (P)
				.1875	.5616438 (BC)
.925	.45625	.0346563	.1047699	.2509048	.6615952 (N)
				.3393398	.8106061 (P)
				.3125	.64375 (BC)
.975	.6	.0644339	.1507743	.3044879	.8955121 (N)
				.4086708	.8700964 (P)
				.3806452	.8543046 (BC)

```

. rocplot, plotiopts(msymbol(none) lcolor(black))
> scheme(sicolor) rlopts(lcolor(black))
> ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
> xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
> legend( col(1) order(1) label(1 "ROC AUC= 0.83 [0.77;0.90]"))

```

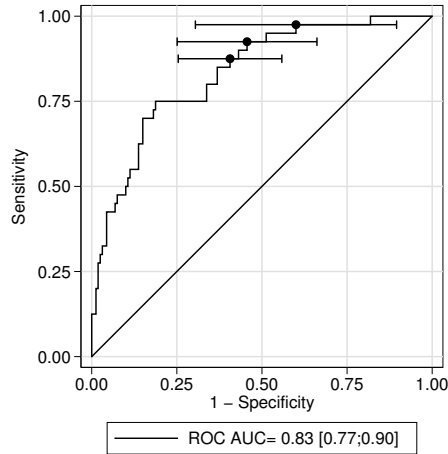


Figure 4. ROC curve for model 1 showing false-positive rate and 95% CI for given true-positive values

The risk thresholds of the prognostic model allow us to see the effects of selecting different cutoff points on medical decisions (Royston, Altman, and Sauerbrei 2006). Further, we can display the ROC curves of different classifiers in the same figure, but we should remember that even when two classifiers agree in sensitivity and specificity, they will not necessarily have the same cutoff point values. Now, we show an example with the classifiers for the first and second models. We use `rocplot` to calculate the sensitivity and 1 – specificity values for both models and then list the predicted probabilities where the sensitivity and 1 – specificity values agree.

```
. rocreg LNM predprob predprob2, nodots
Bootstrap results                Number of obs   =       200
                                Replications     =       1,000
```

```
Nonparametric ROC estimation
Control standardization: empirical
ROC method                : empirical
```

Area under the ROC curve

```
Status      : LNM
Classifier:  predprob
```

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
	.8335938	-.0010747	.0368254	.7614172	.9057703 (N)
				.7531488	.8963632 (P)
				.7459636	.8940851 (BC)

```
Status      : LNM
Classifier:  predprob2
```

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
	.8339062	-.0006906	.0357851	.7637688	.9040437 (N)
				.7543176	.8938994 (P)
				.752333	.891875 (BC)

Ho: All classifiers have equal AUC values.

Ha: At least one classifier has a different AUC value.

P-value: .9741677 Test based on bootstrap (N) assumptions.

```
. list if _roc_predprob==_roc_predprob2 & _fpr_predprob==_fpr_predprob2
```

	LNM	predprob	predpr-2	_roc_p-b	_fpr_p-b	_roc_p-2	_fpr_p-2
12.	0	.002632	.0291194	1	.93125	1	.93125
133.	0	.2548581	.4659569	.75	.2375	.75	.2375

```
. rocreg LNM predprob predprob2, nodots roc(0.2375)
Bootstrap results                Number of obs   =       200
                                Replications     =       1,000
```

```
Nonparametric ROC estimation
Control standardization: empirical
ROC method                : empirical
```

ROC curve

```
Status      : LNM
Classifier:  predprob
```

ROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
.2375	.75	-.0050337	.0736913	.6055677	.8944323 (N)
				.6	.8787879 (P)
				.6136364	.8857143 (BC)

```
Status      : LNM
Classifier: predprob2
```

ROC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]	
.2375	.75	-.0151284	.087794	.5779269	.9220731 (N)
				.5357346	.8837209 (P)
				.5833333	.90625 (BC)

Ho: All classifiers have equal ROC values.

Ha: At least one classifier has a different ROC value.

Test based on bootstrap (N) assumptions.

ROC	P-value
.2375	1

At a sensitivity of 0.75 and 1 – specificity of approximately 0.25, the classifiers agree. We use `rocregplot` to plot the ROC curves of both classifiers and caption the differing cutoff point values (figure 5).

```
. rocregplot, scheme(s1color) rlopts(lcolor(black))
> plot1opts(msymbol(none) lcolor(black))
> plot2opts(msymbol(none) lcolor(black))
> ylabel(0(0.25)1, angle(0) format(%3.2f)) ytitle("Sensitivity")
> xlabel(0(0.25)1, format(%3.2f)) xtitle("1 - Specificity")
> text(0.75 0.19 "25%", size(small) placement(n) color("dknavy"))
> text(0.70 0.29 "47%", size(small) placement(n) color("maroon"))
> legend( col(1) order(1 2)
> label(1 "ROC AUC - Model 1= 0.83 [0.77;0.90]")
> label(2 "ROC AUC - Model 2= 0.83 [0.77;0.90]"))
```

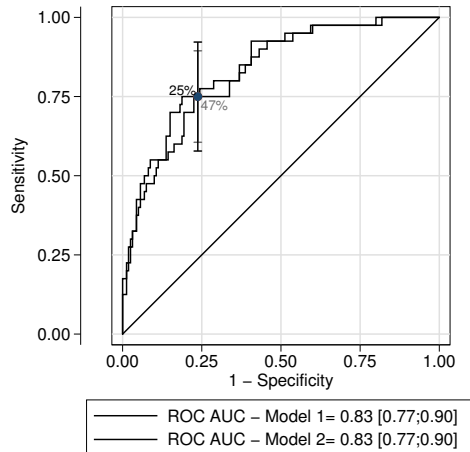


Figure 5. ROC curves for models 1 and 2

In this example, the layout of the thresholds allows us to define what predicted probability value is appropriate to define patients as “low risk” and “nonlow risk” of developing LNM. This will help us to evaluate, for example, the therapeutic value of lymphadenectomy in “nonlow-risk” patients in clinical trials, as in the role of sentinel lymph node biopsy and the prognostic value of metastatic nodal resection in this group and not in the entire population (Kim et al. 2014). That is, it will help us to identify those persons who may benefit from being included in a clinical trial, which lets us better understand disease mechanisms and select the best treatment options.

3 Conclusion

Adding at least one or more informative cutoff points on the ROC graph or textual description of a prognostic model enables one to better characterize the test as well as evaluate the strategic discriminatory capacities. This good practice will help improve our knowledge about the diagnostic and prognostic uses of clinical procedures.

4 References

- DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44: 837–845.
- Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.
- Kim, D.-Y., S.-H. Shim, S.-O. Kim, S.-W. Lee, J.-Y. Park, D.-S. Suh, J.-H. Kim, Y.-M. Kim, Y.-T. Kim, and J.-H. Nam. 2014. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. *British Journal of Cancer* 110: 34–41.
- Moons, K. G. M., D. G. Altman, J. B. Reitsma, J. P. A. Ioannidis, P. Macaskill, E. W. Steyerberg, A. J. Vickers, D. F. Ransohoff, and G. S. Collins. 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine* 162: W1–W73.
- Royston, P., D. G. Altman, and W. Sauerbrei. 2006. Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine* 25: 127–141.
- Zweig, M. H., and G. Campbell. 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39: 561–577.

About the authors

David Lora is a biostatistician at the Department of Clinical Research, Clinical Trials, Clinical Epidemiology and Bioinformatics, Hospital Universitario 12 de Octubre, Madrid, Spain. He is also a researcher at the Spanish Epidemiology and Public Health Consortium. His research interests lie in the intersection of computer science and statistical methods with applications in the areas of prognosis and cancer research.

Israel Contador, PhD, is a professor of psychology at the Department of Basic Psychology, Psychobiology and Methodology of Behavioral Sciences, University of Salamanca, Spain. He is particularly interested in clinical neuropsychology, and his research covers topics such as the prognosis of mild cognitive impairment and the early detection of dementia.

José F. Pérez-Regadera, PhD, is the head of the Department of Radiation Oncology, Hospital Universitario 12 de Octubre, Madrid, Spain. He is an associate professor in the Department of Radiation Oncology, Complutense University, Madrid, Spain. His research focuses primarily on gynaecological cancer, neuro-oncology, and brachytherapy.

Agustín Gómez de la Cámara, PhD, is a senior investigator at and the head of the Department of Clinical Research, Clinical Trials, Clinical Epidemiology and Bioinformatics, Hospital Universitario 12 de Octubre, Madrid, Spain. He is a professor at the Department of Medicine at Madrid University. He is a member of the executive group of the Spanish Clinical Research Network and the European Clinical Research Network and a senior researcher at the Spanish Epidemiology and Public Health Consortium. His research focuses on the efficacy of medical and public health interventions and diagnostic and prognostic models in chronic diseases.