



UNIVERSIDAD NACIONAL DE COLOMBIA

**Desarrollo de un meta-modelo de predicción  
para el diagnóstico de infección cervical por  
*Chlamydia trachomatis* en mujeres de  
Colombia**

**Juan Manuel Reyes S.**

Universidad Nacional de Colombia  
Facultad de Medicina  
Instituto de Investigaciones Clínicas  
Bogotá, Colombia  
2014



**Desarrollo de un meta-modelo de predicción  
para el diagnóstico de infección cervical por  
*Chlamydia trachomatis* en mujeres de  
Colombia**

**Juan Manuel Reyes S.  
Código: 599262**

**Trabajo de grado para optar por el título de Magister en Epidemiología Clínica**

Director:  
Dr. Hernando Gaitán Duarte

Codirector:  
Dr. Alfonso Iorio

Línea de Investigación:  
Modelos de predicción diagnóstica  
Grupo de Investigación:  
Grupo de evaluación de Tecnologías y políticas en salud

Universidad Nacional de Colombia  
Facultad de Medicina  
Instituto de Investigaciones Clínicas  
Bogotá, Colombia  
2014



## Resumen

**Introducción:** La cervicitis por clamidia tiene alta prevalencia, es de difícil diagnóstico (30-75% de los casos son asintomáticos) y pueden tener serias secuelas clínicas. En condiciones de escasos recursos, el diagnóstico y tratamiento principalmente son realizadas bajo el manejo sindrómico propuesto por la OMS, el cual ha mostrado desempeño subóptimo. Se desarrolló y validó un meta-modelo basado sobre una revisión sistemática de modelos predictivos publicados,

**Metodología:** Se condujo una revisión sistemática de reglas de predicción clínica para el diagnóstico de infección de cervicitis por *Chlamydia trachomatis*, buscando en las bases de datos Medline, EMBASE y Bireme. La revisión, selección, extracción de datos y evaluación de la calidad fueron realizadas independientemente por dos investigadores, usando formas estandarizadas y nuevas herramientas específicas para riesgo de sesgo en modelos predictivos. Las características y las medidas de desempeño de los modelos incluidos fueron comparadas y resumidas. Adicionalmente, fueron agregados usando dos métodos disponibles (modelo promediado y regresión de Stacked) y una muestra externa de 1381 mujeres de Colombia con una prevalencia del 9.7% de casos diagnosticados con PCR. La máxima verosimilitud penalizada fue usada para corregir el sobreajustamiento de los coeficientes del meta-modelo obtenidos por la regresión de Stacked y bootstrapping fue usada para realizar la validación interna.

**Resultados:** De 3665 registros identificados, 25 artículos fueron incluidos en la revisión sistemática, en donde se reportaron 31 reglas de predicción diagnóstica. Se encontró que los principales riesgos de sesgos están relacionados con selección de los participantes, tamaño de la muestra y análisis de los datos. Siete modelos fueron agregados. El método de la regresión de Stacked generó un meta-modelo con mejores características sobre los modelos existentes: Área bajo la curva 0.79, puntaje de Brier 0.077 y  $R^2$  0.20. La validación interna mostró escaso sobreajustamiento. La comparación de las razones de verosimilitud positiva y negativa fueron mostraron que el meta-modelo tubo mejor desempeño que el actual algoritmo de la OMS.

**Conclusiones:** La regresión de Stacked permitió agregar modelos de predicción, a través datos de pacientes externos individuales. El meta-modelo desarrollado demostró mejores características operativas que el algoritmo de la OMS, sin embargo es necesario realizar estudios de impacto para evaluar su utilidad clínica.

**Palabras clave:** Cervicitis, *Chlamydia trachomatis*, Modelos de predicción, reglas de predicción clínicas, agregación de modelos, CPR.

## Abstract

**Introduction:** Cervicitis by chlamydia has high prevalence, is difficult to diagnose (30-75% asymptomatic cases) and might have serious clinical sequelae. In resource scarce settings, diagnosis and treatment are based on the WHO's proposed syndromic approach, which has largely suboptimal performance. We developed and validated a meta-model based on a systematic review of published prediction models.

**Method:** We conducted a systematic review on clinical prediction rules (CPRs) for the diagnosis of cervicitis infection by *Chlamydia trachomatis*, searching Medline, EMBASE and Bireme databases. Screening, data extraction and quality assessment were performed independently by two researchers, using standardized forms and a new specific risk of bias tool. The performance characteristics of the included models were summarized, and the models were aggregated using two available methods (model averaging and stacked regression method). and an external sample of 1381 women from Colombia (10% positive cases at PCR). Penalized maximum likelihood was used to correct for overfitting the coefficients of the meta-model obtained by the stacked regression, and bootstrapping was used for further internal validation.

**Results:** Of 3665 records identified, 25 articles reporting on 31 different CPRs were included in the systematic review. The principle risk of bias was related to selection of participants, sample size and data analysis. Seven models were aggregated. The stacked regression method generated a meta-model with improved characteristics over each of the existing models: c-statistic 0.79, Brier score 0.077 and  $R^2$  0.20. Internal validation showed lack of overfitting. Comparison of likelihood ratios for positive and negative tests showed that the meta-model had better operational characteristics than the currently used WHO score.

**Conclusions:** Stacked regression allows aggregation of prediction models, provided that suitable external individual patient data are available. The meta-model demonstrated better performance characteristics than the WHO algorithm, however, a prospective implementation study is needed in order to assess of clinical use.

**Keywords:** Cervicitis, *Chlamydia trachomatis*, prediction models, clinical prediction rules, CPR models aggregation

# Contenido

	Pág.
Resumen .....	V
Lista de figuras.....	IX
Lista de tablas .....	X
Introducción .....	1
<b>1. Descripción del problema.....</b>	<b>3</b>
<b>2. Justificación .....</b>	<b>7</b>
<b>3. Objetivos.....</b>	<b>9</b>
3.1 Objetivo general.....	9
3.2 Objetivos específicos .....	9
<b>4. Marco teórico.....</b>	<b>11</b>
4.1 <i>Chlamydia trachomatis</i> .....	11
4.2 Ciclo de Vida de la <i>Chlamydia Trachomatis</i> .....	11
4.3 Cervicitis .....	11
4.4 Carga de la Enfermedad .....	12
4.5 Pruebas Diagnósticas de Infección <i>Chlamydia trachomatis</i> .....	13
4.6 Factores de Riesgo.....	14
4.7 Factores Predictivos .....	16
4.8 Modelos de predicción .....	18
4.9 Modelos de regresión logística.....	18
4.10 Etapas del desarrollo de un modelo de predicción .....	20
4.11 Construcción del modelo de predicción para diagnóstico.....	20
4.11.1 Tipos de estudio .....	20
4.11.2 Tamaño de la muestra.....	21
4.11.3 Técnicas de selección de variables predictores .....	21
4.11.4 Método por pasos o stepwise .....	22
4.11.5 Manejo de valores perdidos.....	23
4.12 Validación .....	24
4.13 Tipos de medidas en el desempeño de modelos de predicción .....	26
4.13.1 Medidas de desempeño general.....	26
4.13.2 Calibración del modelo de predicción .....	27
4.13.3 Discriminación .....	28
4.14 Actualización.....	28
4.15 Meta análisis de modelos de predicción.....	29

4.15.1 Método de modelo promediado.....	30
4.15.2 Método de la regresión “stacked” .....	31
<b>5. Metodología.....</b>	<b>33</b>
5.1 Revisión sistemática de modelos de predicción clínica para el diagnóstico o tamizaje de cervicitis por <i>Chlamydia trachomatis</i> . .....	33
5.1.1 Definición de la pregunta.....	33
5.1.2 Búsqueda de métodos para la identificación de estudios .....	34
5.1.3 Recolección de los datos .....	34
5.1.4 Evaluación de la metodología de la calidad.....	34
5.1.5 Evaluación del desempeño de los modelos de predicción.....	35
5.1.6 Criterio de selección de los modelos de predicción .....	35
5.2 Agregación de modelos de predicción para el diagnóstico de cervicitis por <i>Chlamydia trachomatis</i> .....	35
5.2.1 Método del modelo promediado .....	35
5.2.2 Método de regresión “stacked” .....	38
5.3 Descripción de la muestra de validación.....	39
<b>6. Resultados .....</b>	<b>41</b>
6.1 Revisión sistemática de modelos predictivos para el diagnóstico o tamizaje de cervicitis por <i>Chlamydia trachomatis</i> . .....	41
6.1.1 Característica de los estudios .....	42
6.1.2 Riesgo de sesgo en los estudios incluidos .....	48
6.1.3 Desempeño de los modelos .....	51
6.1.4 Selección de modelos para la agregación.....	52
6.1.5 Descripción de los modelos incluidos.....	53
6.1.6 Descripción de las variables de los modelos incluidos .....	55
6.2 Agregación de modelos de predicción para el diagnóstico de cervicitis por <i>Chlamydia trachomatis</i> .....	58
6.2.1 Descripción de la muestra externa de validación .....	58
6.2.2 Validación y actualización de los modelos seleccionados .....	59
6.2.3 Método de modelo promediado.....	73
6.2.4 Método de regresión “stacked” .....	74
6.2.5 Análisis de sensibilidad .....	79
<b>7. Discusión.....</b>	<b>81</b>
<b>8. Conclusión .....</b>	<b>88</b>
<b>Bibliografía .....</b>	<b>89</b>

## Lista de figuras

	Pág.
<b>Figura 1</b> Diagrama de flujo prisma de la selección de estudios.....	42
<b>Figura 2</b> Evaluación de la calidad metodológica de los estudios incluidos. (El significado de los colores son: Verde: bajo riesgo de sesgo; amarillo: no es claro; rojo: alto riesgo de sesgo) .....	50
<b>Figura 3</b> Comparación entre las gráficas de calibración de los modelos antes y después de la calibración. ....	66
<b>Figura 4</b> Meta-modelo sugerido por el método de la regresión stacked expresado en ecuación.....	75
<b>Figura 5</b> Curva de calibración de bootstrap para el meta-modelo sugerido por el método de regresión stacked. 1000 replicaciones fueron usadas .....	76
<b>Figura 6</b> Curva ROC para el meta-modelo del método de regresión de Stacked.....	77

## Lista de tablas

	Pág.
<b>Tabla 1</b> Ejemplos de factores de riesgo asociados a cervicitis por <i>Chlamydia trachomatis</i> de acuerdo a estudios encontrados. ....	15
<b>Tabla 2</b> Ejemplos de estudios que determinan factores predictores de infección por <i>Chlamydia trachomatis</i> .....	17
<b>Tabla 3</b> Métodos de actualización de modelos de predicción con desenlace binario(6)..	29
<b>Tabla 4</b> Localización y descripción de la población en los estudios incluidos de la revisión sistemática. ....	43
<b>Tabla 5</b> Objetivos de los estudios incluidos y tipo de predicción clínica. ....	44
<b>Tabla 6</b> Reporte de manejo de predictores en los estudios incluidos en la revisión sistemática.....	45
<b>Tabla 7</b> Tamaño de la muestra de los estudios incluidos en la revisión sistemática .....	46
<b>Tabla 8</b> Criterio y método de selección de variables de los artículos incluidos en la revisión sistemática.....	46
<b>Tabla 9</b> Descripción de la validación de los reglas de predicción en los estudios incluidos en la revisión sistemática. ....	47
<b>Tabla 10</b> Variables de predicción frecuentes en las reglas de predicción incluidas.....	48
<b>Tabla 11</b> Desempeño de las reglas de predicción clínica en estudios incluidos en la revisión sistemática.....	51
<b>Tabla 12</b> Características de los modelos incluidos en los procesos de agregación. ....	53
<b>Tabla 13</b> Descripción de las variables de cada modelo seleccionado para el proceso de agregación .....	56
<b>Tabla 14</b> Características de la muestra externa en las características demográficas, comportamiento sexual y síntomas. ....	58
<b>Tabla 15</b> Cambio de valores de coeficientes durante validación y actualización de los modelos incluidos. ....	60
<b>Tabla 16</b> Habilidad discriminativa y medidas de desempeño durante validación y actualización de modelos seleccionados.....	63
<b>Tabla 17</b> Resultados de pesos para cada modelo por el método de modelo promediado. ....	73
<b>Tabla 18</b> Variables y coeficientes para el meta-modelo por el método ponderado.....	73
<b>Tabla 19</b> Meta-modelo sugerido por la regresión stacked.....	74
<b>Tabla 20</b> Validación interna del Meta-modelo para el método de bootstrapping con 1000 replicaciones.....	75

---

<b>Tabla 21</b> Características operativas del meta-modelo, el manejo sindrómico y prueba rápida para la predicción de infección cervical por <i>Chlamydia trachomatis</i> con diferentes puntos de corte.....	78
<b>Tabla 22</b> Meta-modelo desarrollado por la regresión stacked con la muestra de validación sin mujeres trabajadoras sexuales. ....	79



# Introducción

Las infecciones de transmisión sexual (ITS) son infecciones producidas por bacterias, virus o parásitos en el cual, la transmisión es principalmente por contacto sexual como el sexo vaginal, anal y oral (1). Estas son las mayores causas de enfermedades agudas, infertilidad, incapacidad prolongada y muerte con consecuencias en mujeres, hombres y niños(2). De acuerdo a la organización mundial de la salud (OMS), se estima que 1 millón de personas en el mundo adquiere una infección de transmisión sexual y que anualmente 500 millones de personas se infectan con alguno de los agentes de Chlamydia, gonorrhoea, sífilis y tricomoniasis (1).

Las ITSs producidas por *Chlamydia trachomatis* son una de las infecciones más destacadas y complejas de controlar debido a las características que presenta. Principalmente, entre el 70- 75% de las infecciones por clamidia son asintomáticas (3). Por lo tanto, gran parte de las pacientes no recurren a atención médica, trayendo como consecuencia la continua transmisión de la enfermedad, complicaciones de la infección como enfermedad pélvica inflamatoria, embarazo ectópico y secuelas tales como infertilidad y dolor crónico, (3). De acuerdo al boletín del 2008 de la OMS, las infecciones causada por clamidia estuvieron dentro de los tres ITS con más nuevos casos reportados, según este boletín solamente para clamidia se estimó 105.7 millones de casos(2).

Existen varias pruebas diagnósticas para la determinación de infección por clamidia. En países emergentes y en vía de desarrollado, se utiliza principalmente el diagnóstico sindrómico establecido por la OMS, el cual se basa bajo el hallazgo de factores de riesgo, signos y síntomas de los pacientes (4). Sin embargo, el diagnóstico sindrómico ha presentado bajas características operativas como han mostrado diferentes estudios principalmente por la historia natural de la enfermedad (3,5).

En países desarrollados principalmente pero también emergentes se utilizan las pruebas rápidas que se basan principalmente en pruebas inmunoenzimáticas. Adicionalmente existen otra clase de pruebas relacionadas con inmunofluorescencia (6).

Actualmente, el patrón de oro en esta infección es la amplificación de ácidos nucleicos como el LCR y el PCR. A pesar de que se cuenta con una alta variedad de pruebas diagnósticas, hay factores como el costo, la accesibilidad y las características operativas de las mismas que no han permitido disponer de adecuadas pruebas para la población en general, en especial en países en vía de desarrollo (6).

Gracias a los avances que se han realizado en análisis multivariable, se han planteado diferentes modelos de predicción con el objetivo de mejorar el desempeño del manejo sintromico. Los modelos de predicción presentan la ventaja de establecer probabilidad que una persona este enferma siendo de bajo costo, accesible y rápido.

Los modelos de predicción parten de la agrupación de las variables que presentan los pacientes con la enfermedad diferenciándose de aquellos sin ella (7). Se han desarrollado varios modelos para el diagnóstico de cervicitis por clamidia con regular desempeño. Éstos han tenido diferentes variables de las cuales se destacan friabilidad cervical, mucopus, flujo vaginal, edad, más de 1 compañero sexual entre los últimos 12 meses, nuevo compañero sexual en los últimos meses, compañero sexual sintomático entre otros (8). Adicionalmente, las poblaciones han sido específicas por sus características geográficas o demográficas que limitan su generalización en otras poblaciones.

El presente trabajo tiene por objetivo desarrollar un nuevo modelo de predicción para el diagnóstico de cervicitis producido por *Chlamydia trachomatis* en mujeres Colombianas a partir de previos modelos de predicción desarrollados. Para cumplir este objetivo se partió de una revisión sistemática de modelos de predicción publicados, los cuales fueron evaluados en cuanto a su desarrollo y desempeño. Posteriormente, los modelos seleccionados fueron validados externamente para evaluar su comportamiento en la población colombiana y meta-analizados a través de dos métodos desarrollados recientemente.

# 1.Descripción del problema

Las infecciones del tracto genital como su mismo nombre lo indica, son infecciones del aparato reproductor que se puede dividir en tres tipos: infección endógena que es producida por un crecimiento excesivo de microorganismos que se encuentran normalmente en la vagina; infección iatrogénica cuando se produce una infección a través de un procedimiento médico y finalmente las infecciones de transmisión sexual que son causados por virus, bacterias o parásitos que son transmitidos a través de la actividad sexual con una pareja infectada (9). Estas infecciones son de alta prevalencia tal como lo muestra el reporta de la OMS, en donde se estimó un total de nuevos casos para el 2008 de 498,9 millones de casos curables de *Treponema pallidum* (sífilis), e infecciones por *Neisseria gonorrhoeae*, *Chlamydia trachomatis* y *Trichomonas vaginalis*, entre hombres y mujeres de edades comprendidas entre los 15 y los 49 años. La mayor proporción de casos se presentó en América y el oeste pacifico (2).

Centrándose en *Chlamydia trachomatis*, se encuentra que es uno de los agentes infecciosos más frecuentes dentro del grupo de las infecciones de transmisión sexual. De acuerdo a la OMS para el año 1999 se estimó 13,5 millones de nuevos casos de infecciones por *Chlamydia trachomatis* en el continente americano (3). Para el 2008, la OMS reportó 26,4 millones de nuevos casos en esta misma región. Siendo la segunda infección con más alta incidencia después de *T. vaginalis*, comportamiento que se ha mantenido a lo largo de éste periodo(2).

En la Unión Europea en el 2011 se reportó una prevalencia para la infección clamidial de 175 por 100.000 habitantes (10). En Estados Unidos, el CDC estimó una prevalencia de 456.7 casos por 100000 habitantes (11). Para el contexto Colombiano no se conoce una prevalencia de la enfermedad a nivel nacional debido a que no se disponen estudios que abarquen este alcance de acuerdo a la búsqueda realizada en páginas como Embase, Medline y Lilacs. Sin embargo, se han publicado algunos estudios realizados a nivel de la ciudad capital. De acuerdo al más reciente de éstos estudios realizado en el 2010 se estimó una prevalencia de 9,7%(12).

Por otro parte, la cervicitis por *Chlamydia trachomatis* es una infección de gran importancia por las complicaciones y secuelas que puede traer a largo término. En primer lugar, aumenta el riesgo a desarrollar la enfermedad pélvica inflamatoria (EPI) y sus secuelas como la infertilidad tubular, el embarazo ectópico y el dolor pélvico crónico (13). La relación entre la enfermedad pélvica inflamatoria y la infección por *Chlamydia trachomatis* no ha sido consistente entre los estudios, sin embargo, se han encontrado estudios clínicos que

demuestran reducción en el desarrollo de EPI cuando se realiza tamizaje para infecciones por *Chlamydia* en forma preventiva (14). Adicionalmente, en mujeres embarazadas y con infecciones por esta bacteria se incrementa el riesgo de ruptura prematura de membranas, y de tener infantes de bajo peso(15,16). Los niños nacidos de madres con infecciones por *Chlamydia trachomatis* sin tratar, presentan alto riesgo de desarrollar conjuntivitis neonatal y/o neumonía(17,18). Por otro lado, en las mujeres con infecciones por *Chlamydia trachomatis* son asociadas con un incremento en el riesgo de adquirir infecciones de Virus de Inmunodeficiencia humana (VIH) (19). Debido a las series consecuencias en cuanto a complicaciones y secuelas que genera esta infección es determinante el diagnóstico oportuno.

Sin embargo, el proceso diagnóstico de esta infección es complejo debido a la presentación clínica de cervicitis, en donde el 70-75% de las pacientes son asintomáticas, (3) y la falta del cierre del proceso diagnóstico – tratamiento por el no retorno de las pacientes por los resultados de exámenes diagnósticos y tratamiento (20–22). A pesar que se cuenta con diversas pruebas diagnósticas como el diagnóstico sindrómico, el cultivo y los métodos de caracterización de *Chlamydia trachomatis* diferentes al cultivo (6), la OMS ha optado por el diagnóstico sindrómico en busca de ofrecer una prueba diagnóstica asequible en el punto de atención, de bajo costo, e inicio de tratamiento temprano en sitios de atención donde es limitado el acceso a las pruebas diagnósticas con adecuado desempeño (21).

Se han hecho algunos estudios que miden el impacto del diagnóstico sindrómico en la reducción de la prevalencia de ITS mostrando buenos resultados. El estudio de Kamali (23) se observó disminución de la incidencia de *N. gonorrhoea* RR 0.28 (0.11-0.70) y sífilis con RR 0.58 (0.35 – 0.96) al comparar la efectividad del diagnóstico sindrómico con otras estrategias comunitarias.

Desafortunadamente, las características operativas del diagnóstico sindrómico para el síndrome de cervicitis han sido pobres cuando se utiliza en otras poblaciones como es el caso de Colombia en donde la sensibilidad fue de 13.7% (IC 95% 8.57% – 20.4%) y especificidad de 91% (IC 95% 89.3%-92.5%) (5). Otros estudios han mostrado que la sensibilidad, la especificidad, y la razón de verosimilitud positiva (LR(+)) y negativa (LR (-)) del diagnóstico sindrómico pueden variar entre 30-80%, 40-80%, 1,5 – 4 y 0,87 – 0,25 respectivamente (20,24,25) . Por lo tanto, el comportamiento de estas características operativas pueden variar dependiendo del grupo de población estudiada. Esto puede observarse en el estudio realizado en Oeste de África, en donde la prevalencia de infección por *Chlamydia trachomatis* fue de 3,2 % presentando una sensibilidad y valor predictivo positivo del 60% y 5% (26) pero en otra población de África con una prevalencia de 24,5% se obtuvo un sensibilidad del 48,3% y valor predictivo positivo de 38,3% (20).

Ante este escenario, se ha intentado mejorar la sensibilidad y la especificidad del diagnóstico sindrómico involucrando en el algoritmo variables correspondientes a características demográficas (edad, raza, estado civil, etc), signos y síntomas (características del flujo

vaginal) y comportamientos sexuales de riesgo (número de compañeros sexuales en los últimos 12 meses, nuevos compañeros sexuales, entre otros) (20,27,28).

Por esta razón, a través de herramientas estadísticas, se han desarrollado modelos de predicción en la cual se combinan factores predictivos con el objetivo de clasificar mejor los pacientes (7). Focalizándose en cervicitis, varios modelos de predicción han presentado mejores características operativas en comparación al diagnóstico sindrómico (29–31) como por ejemplo el modelo de Ruggao en donde la sensibilidad y especificidad fue 72.8% y 59.8% respectivamente (28). Sin embargo, es necesario tener cuidado al momento de analizar e implementar un modelo predicción debido a que hay varios factores que pueden generar sesgos en los resultados como son el tamaño de la muestra, método de selección de las variables, medición de los factores de predicción y el desenlace, entre otros (32).

Para mejorar el desempeño de los modelos de predicción, se han desarrollado varias técnicas estadísticas y criterios metodológicos(32). Entre las técnicas más destacables se encuentran las técnicas avanzadas de validación, revisión y actualización que buscan mejorar el ajuste de un modelo previamente desarrollado a una muestra externa (33). Éstas ofrecen muchas ventajas, pero por su reciente desarrollo no han sido aplicadas a los modelos de predicción diagnóstica de cervicitis.

Adicionalmente, teniendo en cuenta las limitaciones en tamaño de la muestra y los diversos factores de predicción que se requieren para desarrollar un modelo de predicción diagnóstica, una nueva técnica de meta-análisis de modelos de predicción fue desarrollada buscando crear un modelo con mejores posibilidades de generalización y mejores características operativas a partir de modelos ya publicados (34).

De otra parte, tanto en Latinoamérica como en Colombia no se encuentran estudios que involucren el desarrollo de reglas de predicción clínica para el diagnóstico de infección por *Chlamydia trachomatis*. Generalmente, estos estudios que se refieren a esta infección se encuentran más relacionados hacia determinación de factores de riesgo pero en poblaciones de alto riesgo.

Por lo tanto, podemos evidenciar varios problemas en la relevancia y diagnóstico oportuno de la cervicitis por *Chlamydia trachomatis*, las pobres características operativas del diagnóstico sindrómico, la deficiencia de los modelos predictivos desarrollados para la población colombiana, la falta de aplicación de técnicas avanzadas en desarrollo y la validación de modelos predicción diagnóstica de esta infección y la falta de reglas de predicción diagnóstica en la población colombiana. Este trabajo se enfocará en el mejoramiento del diagnóstico de cervicitis por *Chlamydia trachomatis* a través de un meta-modelo de predicción diagnóstica en mujeres con cervicitis de Colombia a partir de modelos de predicción desarrollados en otras poblaciones.



## 2. Justificación

Como ya se ha planteado, existen varios problemas en el diagnóstico de la infección cervical por *Chlamydia trachomatis*, en los cuales al no tener un diagnóstico temprano pueden producirse importantes consecuencias para las pacientes debido a las complicaciones y secuelas de la enfermedad. Igualmente, se observa aumento en la incidencia de la enfermedad por el continuo ciclo de transmisión de la infección en la población. De aquí radica la importancia de esta investigación en donde se busca mejorar el diagnóstico de la infección de la cérvix producida por *Chlamydia trachomatis* a través de modelos de predicción.

El impacto de desarrollar un modelo de predicción con características operativas adecuadas y con buen desempeño en la población colombiana para el diagnóstico de cervicitis sería alto desde muchas perspectivas. Desde el punto de vista de epidemiológico, disminuiría la carga de la enfermedad dado que se cortaría el ciclo de transmisión y habría detección temprana produciendo disminución de pacientes con secuelas y reinfecciones. Adicionalmente, el ahorro en costos en salud por disminución de la atención médica y todos los recursos que este conlleva. Y desde el punto de vista clínico, sería una herramienta importante para el ginecólogo o médico general que le daría mayor certeza al diagnóstico e implementación del tratamiento médico.

Por otra parte, en varios estudios se ha evidenciado que el comportamiento de los factores de riesgo y factores predictores pueden ser diferentes en cada grupo social, ya sea por la raza, la prevalencia de la enfermedad, los comportamientos sexuales, los estilos de vida, el nivel socio-económico, entre otros. Partiendo de este hecho, se necesita trabajar con las características propias de la sociedad colombiana para crear un modelo que permita incrementar la probabilidad diagnóstica de infección por *Chlamydia trachomatis* propias de la población.

Adicionalmente, este trabajo permite la generación y ampliación del conocimiento sobre modelos de predicción puesto que se estaría trabajando más allá del desarrollo como la validación externa, la actualización y la agregación de modelos que son tópicos que se han profundizado ampliamente. Así mismo, este trabajo sería uno de los pocos realizados en este campo a nivel nacional.

La agregación de modelos de predicción principalmente han sido aplicados en campos relacionados con desordenes sanguíneos en donde la patología y los factores predictivos han mostrado mayor correlación en comparación con la cervicitis. Este estudio demostraría el uso de un nuevo método de agregación de modelos de predicción.



## **3. Objetivos**

### **3.1 Objetivo general**

Desarrollar un modelo predictivo para el diagnóstico de infección cervical por *Chlamydia trachomatis* basado en previos reportes de modelos de predicción y las características de las mujeres de Bogotá.

### **3.2 Objetivos específicos**

1. Evaluar los modelos predictivos para el diagnóstico de infección cervical por *Chlamydia trachomatis* reportados en la literatura.
2. Validar los modelos predictivos seleccionados en la literatura en la población de mujeres de Bogotá.
3. Actualizar los modelos de predictivos para el diagnóstico de infección cervical por *Chlamydia trachomatis* a la población de mujeres de Bogotá.
4. Meta-analizar los modelos encontrados en la literatura para el diagnóstico de infección cervical por *Chlamydia trachomatis*.



## 4. Marco teórico

### 4.1 *Chlamydia trachomatis*

La especie Chlamydia es una bacteria aerobio obligados y patógenos intracelular debido a su incapacidad de producir ATP utilizando células huésped como fuentes de producción de energía (35).

Existen varios serotipos de *Chlamydia trachomatis* de acuerdo a las proteínas de membrana. Se han detectado 15 serotipos: A, B, B<sub>a</sub>, C, D, E, F, G, H, I, J, K, L<sub>1</sub>, L<sub>2</sub> y L<sub>3</sub>. Los serotipos D a la K son los responsables de infecciones oftálmicas, genitales y rectales llevando a conjuntivitis, cervicitis y proctitis.

### 4.2 Ciclo de Vida de la *Chlamydia Trachomatis*

El ciclo de la vida de la Chlamydia consiste en dos distintas fases. En la fase infecciosa se encuentra como cuerpos elementales, el cual busca unirse a la pared celular de las células para inducir su fagocitosis. Las células que ataca principalmente son epiteliales columnares o cúbicas las cuales se encuentran en la conjuntiva, uretra, endocervix y mucosa del endometrio y trompas de falopio. En su fase de replicación, forma el estado de cuerpo reticular en donde es activo metabólicamente y forma nuevos cuerpos elementales que posteriormente son liberados. El ciclo de infección tiene un tiempo aproximado de 48-72 horas (36). El proceso de incubación es de 7-21 días(35).

### 4.3 Cervicitis

La cervicitis es una inflamación de los tejidos que recubren el cuello uterino que puede ser de origen infeccioso o no infeccioso. En la cervicitis de origen infeccioso puede ser causado por *Chlamydia trachomatis*, *Neisseria Gonorrhoeae*, virus herpes simple o combinaciones de estos. Este tipo de cervicitis se transmite a través del acto sexual o durante el parto.

La cervicitis causa tres principales complicaciones: propagación intraluminal ascendente de microorganismos patógenos que producen enfermedad pélvica inflamatoria, infección ascendente en embarazo produciendo ruptura prematura de embarazo, infección puerperal y neonatal, y finalmente desarrollo de neoplasia cervical (37).

## 4.4 Carga de la Enfermedad

Las infecciones por *Chlamydia trachomatis* es una de las bacterias más comunes en las enfermedades de transmisión sexual. En el panorama mundial, de acuerdo al reporte de la OMS en el 2008 se estimó una incidencia de 105.7 millones de casos por año ocupando el segundo lugar más frecuente después de *T. vaginalis* con 276.4 millones de casos por año. Al mirar por distribución geográfica, se reportaron 40 millones de nuevos casos en el Oeste pacífico, 26,4 millones en el continente americano, 20,6 millones en Europa, 8.3 millones en el continente africano, 7.2 millones en Sur-este de Asia y 3.2 millones en el este mediterráneo. Éste agente infeccioso comparado con los demás infecciones fue el segundo más frecuente en las regiones de América, Europa y el Este del mediterráneo. En los otros regiones del mundo estuvo en tercer lugar por debajo de *T. Vaginalis* y *N. gonorrhoeae* (2).

En Estados Unidos, el CDC para el 2012 reportó un total de 1´422.976 de casos donde se estimó una prevalencia de 456.7 casos por 100.000 habitantes mostrando que es una de los infecciones más frecuentes en el país (11). Éste mismo comportamiento se fue observado en Europa de acuerdo a los reportes de la European Centre for Disease Prevention and Control (ECDC)(10).

Sin embargo, se estima que estos reportes son subregistros de las cifras reales de *Chlamydia trachomatis*, sobre todo en las poblaciones de mayor riesgo. Siendo la principal causa de esta situación la ausencia de síntomas en la población infectada como por ejemplo en el caso se Estados Unidos se ha estimado que solo se reporta entre el 50 – 80% de los casos al sistema de vigilancia (38).

Las infecciones producidas por *Chlamydia trachomatis* son de gran importancia desde la visión de salud pública debido a la alta carga de enfermedad y su forma de crecimiento de la misma. En Estados Unidos, desde 1992 a 2012 la tasa de infección incremento de 182.3 casos a 456.7 casos por 100.000 habitantes. Este incremento es mayor en mujeres en comparación de los hombres. En las mujeres sexualmente activas entre las edades de 16-24 años para el 2001, el incremento fue de 23% mientras que para el 2011 fue 45,1% (11). Sin embargo, en los últimos años este crecimiento ha disminuido tanto en Europa con en los Estados Unidos (10,11).

La distribución en género y edad de la infección por *C. trachomatis* es más alta en mujeres que en hombres y entre las edades de 15-24 años (10,11). Sin embargo, hay que tener en cuenta el posible sesgo de selección debido a que esta información proviene de países desarrollados en donde las poblaciones son altamente tamizadas.

En Colombia, se han realizado 3 estudios en Bogotá de corte transversal detectando casos prevalentes de la enfermedad. Uno de estos estudios fue realizado tanto en

mujeres sintomáticas como asintomáticas encontrando una prevalencia de 5,6% (IC 95% 3,7-9,1)(39). Por otro lado, el estudio de Franceschi el cual realizó muestras probabilísticas por edades encontró una prevalencia de 7,2 IC 95%(4,4 – 10,9)(40). En el estudio más reciente realizado en mujeres sintomáticas encontró una prevalencia de 9,7% (12).

La prevalencia a nivel nacional se desconoce por falta de estudios de acuerdo a la búsqueda realizada en Lilacs, Embase y Medline.

## 4.5 Pruebas Diagnósticas de Infección *Chlamydia trachomatis*

El cultivo fue considerado el patrón de oro en el siglo XX para el diagnóstico de infección por clamidia. Pero con la llegada de las pruebas amplificadoras de ácidos nucleicos (reacción en cadena de polimerasa (PCR) y reacción en cadena de ligasa (LCR)) el uso de ésta prueba diagnóstica ha decaído. Estudios realizados en comparación con PRC, el cultivo presentó una sensibilidad de 70 – 90%, especificidad del 100% y LR(+) 0,32-0,1 (41). Por otro lado, presenta varias desventajas dado que los especímenes requieren ser transportados en cadena de frío, es costosos, requiere entre 3-7 días para obtención de resultados y necesita un técnico experto en el proceso (36).

Las pruebas rápidas que son basadas en la detección de antígenos presentan una sensibilidad relacionado con el cultivo de 43-92% y especificidad de 92-100% (42). Presentan el problema que su desempeño depende de la prevalencia de la enfermedad, como consecuencia el valor predictivo positivo es muy variable. Es fácil de usar y el resultado se obtiene en menos de 30 min (43).

También se cuenta con pruebas serológicas que consisten en cuantificar los marcadores serológicos en sangre como son las inmunoglobulinas IgG, IgA e IgM específicos debido al incremento en el proceso infeccioso con *Chlamydia trachomatis*. Se cuenta con dos pruebas: la microinmunofluorescencia y la prueba ELISA. Presentan el problema de tener reacción cruzada con *Chlamydia pneumoniae* afectando la especificidad de la prueba (44). Para el marcador de IgG utilizando ELISA, la sensibilidad varía de 60 - 86% y la especificidad es de 30 – 70% (45–47).

Actualmente, se considera como el patrón de oro las pruebas de amplificación de ácidos nucleicos por su alta sensibilidad y especificidad. Entre estas pruebas se tiene la reacción en cadena de la polimerasa (PCR) y reacción en cadena de ligasa (LCR) que tienen una sensibilidad de 92-96% y especificidad del 98-100% (48,49). El tiempo de reporte de resultados es mayor a 4 horas requiriendo de personal y equipos sofisticados para su interpretación (43).

Finalmente, el diagnóstico sindrómico propuesto por la OMS tiene las ventajas de tener inicio oportuno de tratamiento y de accesibilidad a la población general. Las características operativas de sensibilidad, especificidad, LR(+) y LR (-) varía entre el 30-80%, 40-80%, 1,5 – 4 y 0,87 – 0,25 respectivamente (20,24,50).

## 4.6 Factores de Riesgo

De acuerdo a la OMS, un factor de riesgo está definido como cualquier atributo, característica o exposición de un individuo que incrementa la probabilidad del desarrollo de una enfermedad (51)

Se han realizado varios estudios en diferentes poblaciones buscando factores de riesgo que se asocien en el desarrollo de las infecciones de transmisión sexual y en especial de *Chlamydia trachomatis* por su alta prevalencia en comparación con las otras infecciones.

El gran problema de estos estudios es el riesgo de sesgo de mala clasificación debido a uso de pruebas diagnósticas con sensibilidad y especificidad variables como son las pruebas de inmunofluorescencia.

Una revisión sistemática realizada en Inglaterra de estudios de prevalencia que se han realizado en ese país, encontró la edad como factor de riesgo con significancia estadística, en donde las edades entre 20 - 24 años se comportan como protectores con un OR de 0,62 (0,52 – 0,75) y mayores a 30 años 0,16 (0,12- 0,22) tomando con referencia basal sujetos con edades menores de 20 años(52).

Otros estudios convergen en muchos factores de riesgo como son el número de compañeros sexuales en el último año, la edad del primer coito, la raza, el no uso de condón y un nuevo compañero sexual(53–55). En el estudio realizado por K W Radcliffe y colaboradores (53) encontró que a menor edad, mayor número de compañeros sexuales en el último año, y el uso infrecuente de condón aumenta el riesgo de infección de *Chlamydia trachomatis* con OR significativos. Otro factor que también algunos estudios encuentran significativo es el uso de anticonceptivos orales, como lo muestra el estudio de Gavin Hart (55) en donde se encontró un OR 1,8 IC 95% (1,4 – 2,3).

En estos estudios también se han analizado otros factores como fumar, consumo de alcohol y el estrato social. Sin embargo, existe una gran variación entre los resultados obtenidos que dependen del diseño epidemiológico y la población analizada. Por ejemplo, en Brasil se realizó un estudio en población de bajos ingresos en donde se encontró que la disponibilidad de condones en la casa fue un factor de riesgo con un valor de OR de 2,2 (1,1– 4,2)(56).

Es importante recalcar que países como Inglaterra, Estados Unidos, Noruega, China y en África Subsahariana, se han realizado estudios para encontrar factores de riesgo asociados a cervicitis. Caso contrario a la situación nacional en donde no se dispone de

estudios que evalúen factores de riesgo de esta infección en población general. Por ende, se desconoce si los factores de riesgo encontrados en otros países son extrapolables a la población Colombiana o haya otros factores que sean determinantes para esta población.

La tabla a continuación relaciona algunos estudios que tenían por objetivo estudiar factores de riesgo asociados a la infección de *Chlamydia trachomatis*.

**Tabla 1** Ejemplos de factores de riesgo asociados a cervicitis por *Chlamydia trachomatis* de acuerdo a estudios encontrados.

Estudio	Diseño (Lugar)	Factores de Riesgo	Sesgos
Adams y Charlett et al, 2004 (52)	Revisión Sistemática (Inglaterra)	Edad: se estudió entre, 20-24, 25-29 y >30 años. Se obtuvo usando como riesgo basal menor a 20 años OR de 0,62 (0,52-0,75), 0,30(0,23-0,39) y 0,16(0,12-0,22) respectivamente	Se utilizaron diferentes tipos de pruebas diagnósticas en los diferentes estudios
Radcliffe, Ahmad et al. 2001 (53)	Casos y Controles (Reino Unido)	Se encontraron 5 factores de riesgo significativos estadísticamente. Edad: < 20 años (Riesgo basal), 20- 24 OR 0,4 IC (0,3-0,6)), 25-30 OR 0,4 IC (0,3-0,6), > 30 años OR 0,2 IC (0,2-0,4). Total de Numero Compañeros Sexuales: 0 OR 0,1 IC (0,02-1,05), 1 compañero OR 0,4 IC(0,3-0,6), 2 compañeros OR 0,7 IC (0,5-0,9) y > 2 compañeros usado de referencia Uso de Condón: Nunca (Riesgo basal) , Siempre OR 0,4 IC (0,3-0,7) y Algunas Veces OR 0,8 IC (0,6-1,1) Estado Marital: Soltero OR 1,8 IC(1,1-3,1), Casado(Riesgo basal) y Otro OR 0,7 IC(0,3-1,6) Grupo Étnico: Blanco (Riesgo basal), Negro Caribe OR 2 IC (1,5-2,7) y Otro OR 1,2 IC (0,9-1,9)	Pérdidas del 17 % y uso de prueba Diagnóstica con ELISA que tiene menor sensibilidad, sesgo de mala clasificación
Hart. 1992 (55)	Corte Transversal (Australia)	Se encontraron 4 factores de riesgo significativos estadísticamente Contacto ITS OR 4,3 IC (3,3 – 5.6) Uso Anticoncepción oral OR 1,8 IC (1,4 – 2,3) Edad < 25 años OR 3,3 IC ( 2,5 - 4,4) No estar casado o formalmente casado OR 2,0 IC (1.3 – 3.8)	Ellos evaluaban episodios de <i>Chlamydia trachomatis</i> , donde 28% no se les hizo evaluación de chlamydia por no presentar en el primer episodio. Uso de EIA como prueba, posible sesgo de mala clasificación.
LaMontagne,	Corte	Se encuentra 4 factores de riesgo	Posible sesgo de

Estudio	Diseño (Lugar)	Factores de Riesgo	Sesgos
Fenton et al. 2004(57)	Transversal (Inglaterra)	significativos estadísticamente. Edad: Menor 16 años (riesgo basal), 16-19 años OR 1,85 IC(1,41 – 2,42), 20- 24 años OR1,38 IC(1,04 – 1,82) Grupo Étnico: Blanco (Riesgo Basal), Negro Caribe OR 2,04 IC (1,50 – 2,76), Negro Británico OR 1,77 IC (1,28 – 2,46) Nuevo Compañero sexual OR 1,32 IC (1,14 – 1,53) 2 o más compañeros sexual OR 1,57 IC (1,35 – 1,82).	mala clasificación por uso de diferentes pruebas de diagnóstica
Imai, Nakao et al. 2010(58)	Corte Transversal (Japon)	Historia de al menos 4 compañeros sexuales OR 3,17 IC (2,44-4.12), Uso inconsistente de Condón OR 2,74 IC (1,91-3.92) En los últimos seis meses al menos 2 compañeros OR 1,62 IC(1.26-2.10). Edad de su primer Coito: <14 años (Riesgo basal), 18-19 años OR 0,48 IC(0,26 – 0,90), 20 – 21 años OR 0,38 IC (0,18-0,81), mayor 22 años OR 0,23 IC(0,05 – 0,90)	No se estudió población menor a 18 años que tienen mayor riesgo de Chlamydia. Sesgo selección.
Miranda, Szwarcwald. 2004 (56)	Corte Transversal (Brasil)	Edad: 15- 16 años OR 0,4 IC ( 0,3-0.9), 17-19 años. No tener condón en casa OR 2,2 IC (1,1 – 4.2), > 1 compañero sexual en los últimos 12 meses OR 4,7 IC (2,3 – 9,5), No uso de condón regularmente OR 2,9 IC (1,1 – 8,6), Antecedentes de ITS OR 2,3 IC(1,1-5,3)	Se seleccionó una población de bajos recursos. Sesgo de selección. De la muestra seleccionada el 30 % fue sexualmente no activa. Afectando el poder del estudio

## 4.7 Factores Predictivos

Los factores predictivos son atributos o características que permiten establecer la probabilidad de una persona tenga o vaya a desarrollar la enfermedad, a su vez se pueden comportarse como factores de riesgo.

Los factores predictivos se pueden clasificar en (7):

- Demográficos (Edad, Sexo, raza, etc.)
- Tipo y severidad de la enfermedad (diagnóstico principal, características de la enfermedad, síntomas)
- Antecedentes ( Previos episodios, factores de riesgo)
- Comorbilidades
- Estatus funcional físico (Escala de la OMS)
- Calidad de vida y estado de salud subjetivo ( psicológico, función psicosocial)

En la cervicitis por clamidia, los factores predictivos van encaminados a características demográficas como la edad y la raza; antecedentes de riesgo como números de compañeros sexuales en el último 3 meses o el año, nuevo compañero sexual en los últimos tres meses, y previas ITS; y clínicos como mucopus, cervicitis, friabilidad cervical, dolor abdominal, flujo vaginal entre otros (27,28,59,60). En la tabla 2 se muestran algunos ejemplos.

**Tabla 2** Ejemplos de estudios que determinan factores predictores de infección por *Chlamydia trachomatis*

Estudio	Diseño (Lugar)	Factores de Predictivos
Rugpao, Rungruengthanakit, et al. 2010 (28)	Estudio de Cohorte (Tailandia)	Edad: < 25 años HR 1.7 IC (1.3–2.2) > 25 años (Riesgo Basal) Estado Marital: Poligamia HR 2.0 IC (1.2–3.3) Tener un nuevo compañero sexual HR 3.8 IC (2.7–5.5) Flujo Vaginal Anormal HR 2.1 IC(1.3–3.4) Color del Flujo del mucopus cervical: Claro (Referencia), Blanco/Cremoso/Gris HR1.5 IC (1.1–2.0), Amarillo/Gris HR 3.1 IC (1.3–7.7) Sangrado severo en friabilidad cervical HR 2.6 IC(1.2–5.7)
Tyndal, Kidula et al (27)	Corte Transversal (Kenya)	Edad: <25 años OR 1,73 IC (1,12 - 2,69), > 25 años (Referencia Basal) Más de 2 compañeros sexuales en el último años OR 2,00 IC (1,10 - 3,64) Ser soltero OR 1,58 IC (1,00 – 2,99) Presencia Flujo Cervical OR 2,39 IC ( 1,43 – 4,00)
Hilger, Smith et al. 2001 (60)	Corte Transversal (Estados Unidos)	Edad: 22-24 años (referencia) 14-17 años OR 2,23 IC ( 1,67-2,98) Cervicitis mucopurulenta amarillo o verde OR 3,42 IC (2,01 - 5,80) Friabilidad Cervical OR 2,16 IC (1,55-3,02) Síntomas (Dolor pélvico) OR 1,84 IC(1,29-2,64) Historia de riesgo (múltiples compañeros sexuales, nuevo compañero sexual y contacto con personas con síntomas de Chlamydia) OR 1,56 IC (1,27 - 1,91) Raza (Negro versus Blanco) OR 1,24 IC (1,05-1.47)
Gershman and Barrow. 1996 (61)	Corte transversal (Estados Unidos)	Edad < 19 años RP 3,84 IC (2,88 - 5,13), 20-24 años RP 2,44 IC (1,80 - 3,29), > 25 años (Referencia) Friabilidad cervical RP 2,26 IC(1,63 - 3,15) Mucopus RP 2,64 IC(1,93-3,61) Más de 1 compañero sexual RP 1,40 IC(1,12 -1,75) Raza Negra RP 2,41 IC (1,69- 3.44) Blanco (Referencia)

La fuerza del predictor ésta determinado por la asociación del predictor con el desenlace y la distribución del mismo (7). Es decir, un predictor puede mostrar mayor fuerza cuando

la asociación es marcada pero a su vez es prevalente dentro del grupo en la población objetivo. Por ejemplo, en el estudio de Ruggao la edad se considera fuertemente relacionada debido a que presentó un HR 1,7 (1.3-2.2) y una prevalencia del 50% en la población (28).

## 4.8 Modelos de predicción

Los modelos de predicción son algoritmos que combinan múltiples predictores para estimar la probabilidad que un desenlace esté presente (diagnóstico) o predecir que ocurra (pronóstico)(62).

En los modelos de predicción se pueden tener con diferentes tipos de desenlaces. Entre estos están desenlaces binarios, ordinales, continuos y de supervivencia (7). En el caso específico de diagnóstico, generalmente se está refiriendo a desenlaces binarios.

Para desenlaces binarios, el modelo de regresión logística es el más común y ampliamente usado como técnica estadística. Esta técnica es flexible porque puede combinar predictores categóricos, continuos, transformaciones no lineales y términos de interacción(7).

## 4.9 Modelos de regresión logística

La regresión logística es una técnica que permite determinar la probabilidad de que ocurra un desenlace binario  $Y$  a partir de la combinación lineal del predictor  $x$ . Éste se realiza bajo la transformación logística (logit) (7). En términos matemáticos se puede expresar (ecuación 2-1):

$$P(Y = 1) = \frac{1}{1 + e^{-(\alpha + \beta * x)}} \quad (2-1)$$

Realizando la transformación logit se tendría:

$$\text{Logit}(P(Y = 1)) = \ln \left[ \frac{P(Y = 1)}{1 - P(Y = 1)} \right] \quad (2-2)$$

Reemplazando con la ecuación 2-1 se obtendría:

$$\text{Logit}(P(Y = 1)) = \alpha + \beta * x = \ln p \quad (2-3)$$

Donde  $\alpha$  es el intercepto,  $\beta$  es el coeficiente de regresión correspondiente a  $x$  y éste último es el predictor.  $\ln p$  corresponde a la línea del predictor.

El significado de  $\beta$  corresponde al aporte del incremento en una unidad del predictor  $x$  (7).

En términos de odds éste se expresaría como:

$$\text{Logit}(P(Y = 1)) = \ln(\text{odds}(P(Y = 1))) = \alpha + \beta * x$$

(2-4)

Debido a que odds es igual

$$\text{Odds}(P(Y = 1)) = \left[ \frac{P(Y = 1)}{1 - P(Y = 1)} \right]$$

(2-5)

Si se supone  $x+\Delta x$  como la variación de  $x$ , la transformación de logit puede calcular la relación de odds (OR) de  $x$  en relación con  $Y$ . Por lo tanto, dado que el OR es definido como:

$$\text{OR} = \frac{\text{odds}(P(Y = 1)|(x = x + \Delta x))}{\text{odds}(P(Y = 1)|(x = x))}$$

(2-6)

Al reemplazar con la ecuación 2-4, se tendría:

$$\text{OR} = \frac{e^{\alpha + \beta * (x + \Delta x)}}{e^{\alpha + \beta * x}} = e^{\beta * \Delta x}$$

(2-7)

Cuando la probabilidad de  $Y$  depende de un solo predictor se refiere a un efecto no ajustado.

En cambio, en el caso de varias variables se habla de un efecto ajustado, en donde, está condicionado al comportamiento de los otros predictores(7).

En el caso que se tenga varias predictores, la probabilidad de  $y$  estaría dado como:

$$P(Y = 1) = \frac{1}{1 + e^{-\alpha + \beta_1 * x_1 + \beta_j * x_j}}$$

(2-8)

En términos de transformada se tiene:

$$\text{Logit}(P(y = 1)) = \alpha + \beta_1 * x_1 + \beta_j * x_j$$

(2-9)

En general los modelos de predicción deben cumplir con los supuestos de linealidad y aditividad. Específicamente hablando de los modelos logísticos, la linealidad asume que la variable  $X$  es lineal a  $(\log [P/(1-P)])$  donde  $P$  es la probabilidad del evento. El supuesto de aditividad dice que el efecto de los predictores es aditivo, en este supuesto entra el concepto de interacción donde hay un efecto multiplicativo entre dos factores predictivos.

Para la determinación de los coeficientes de las variables en regresión logística se utiliza la función de verosimilitud que está basada en los principios de mínimos cuadrados utilizado en la regresión lineal. La función de verosimilitud es obtenida del producto del aporte de cada variable como muestra la siguiente ecuación.

$$L = \prod (Y_i * \ln(p(x_i)) + (Y_i - 1) * (\ln(1 - p(x_i)))) \quad (2-10)$$

Se utiliza el logaritmo de verosimilitud ( $LL$ ) por facilidad matemática donde se obtiene:

$$LL = \sum (Y_i * \ln(p(x_i)) + (Y_i - 1) * (\ln(1 - p(x_i)))) \quad (2-11)$$

## 4.10 Etapas del desarrollo de un modelo de predicción

El desarrollo de modelos de predicción generalmente están dividido en tres pasos generales(62):

Paso 1. *Construcción del modelo de predicción.* En esta etapa el modelo es desarrollado a partir de los predictores candidatos. Se utiliza la información de una muestra de población al cual ha sido medido el desenlace y los diferentes predictores. En esta etapa se estiman los coeficientes de cada predictor, la exactitud de la regla de predicción, el potencial optimismo de la regla usando técnicas de validación interna.

Paso 2. *Validación y actualización.* Evalúa la exactitud del modelo en una población externa diferente a la utilizada para desarrollar el modelo. La validación puede ser geográfica, temporal y de diferentes dominios. El modelo puede actualizarse combinando el modelo desarrollado y los datos de la población externa.

Paso 3. *Evaluación de utilidad clínica del modelo de predicción.* Se determina si la aplicabilidad del modelo en la práctica clínica mejora procesos relevantes clínicos, desenlaces en los pacientes y reducción de costos. La habilidad discriminativa permitirá evaluar su relevancia clínica seleccionando las probabilidades predictivas más sensibles.

## 4.11 Construcción del modelo de predicción para diagnóstico

### 4.11.1 Tipos de estudio

Los diseños epidemiológicos utilizados para el desarrollo de modelos de predicción de diagnóstico pueden ser estudios clínicos aleatorizados, cohortes, corte transversal y casos y controles. Sin embargo, el más común son estudios de corte transversal donde las características predictivas de los pacientes son relacionadas a un diagnóstico (7).

Los estudios de corte transversal parten de una cohorte de pacientes con sospecha de presentar una enfermedad. Estos pueden estar derivados de un estudio de cohorte, pero conceptualmente son corte transversal por que evalúan la relación de los predictores y el desenlace en un mismo punto en el tiempo (7). El diagnóstico debe realizarse con el patrón de oro, el cual debe realizarse sin conocimiento de los predictores con el objetivo de prevenir sesgos de información (63).

El problema en el diseño de este tipo de estudios es la incompleta recolección de la información de todos los predictores debido a que las características no son medidas a todos los pacientes en los diferentes perfiles de riesgo de tener la enfermedad.(7)

Los estudios prospectivos presentan la ventaja de la recolección más completa de la evaluación de los factores de predicción, en donde, se tienen definiciones claras y consistentes de éstos.

El diseño de casos y controles presenta grandes limitaciones principalmente con el sesgo de selección por la representatividad de la población seleccionada para todos los pacientes con el diagnóstico de interés (7).

#### **4.11.2 Tamaño de la muestra**

En los diferentes diseños epidemiológicos el tamaño de la muestra se encuentra principalmente determinado por la proporción del evento de interés, la diferencia en la frecuencia del evento en los grupos comparados, la varianza del evento, el nivel de confianza y el poder del estudio.

En modelos de predicción estos mismos parámetros son aplicados enfocándose en el desenlace binario o en un predictor específico (7). Sin embargo, para obtener predicciones confiables se estableció que ésta debe ser mínimo 10 desenlaces por cada predictor candidato (EPV en su siglas en ingles). Este punto de corte fue definido a partir de simulaciones realizadas a diferentes número de desenlaces por predictores candidatos, en donde, se encontró que cuando se tiene un número de desenlaces inferior a 10 EPV se puede obtener inexactitud de la estimación de los coeficientes, disminución del poder estadístico, incremento de la varianza de la regresión y problemas con la convergencia del modelo (64,65).

Es importante aclarar que los predictores candidatos son todos los considerados como potenciales y no solamente las que son incluidos en el análisis multivariable (32).

#### **4.11.3 Técnicas de selección de variables predictores**

La selección de variables predictores se pueden considerar como una forma de reducir el número de variables debido a que modelos grandes serían menos prácticos y habrían variables predictores con menor efecto o implausibles efectos (7).

En este momento se incluye el concepto de parsimonia en donde se establece que simples explicaciones son preferidas sobre explicaciones más complejas, es decir, que un modelo en donde se retiran del análisis variables con menor significancia sería más parsimonioso(7).

Se han desarrollado varias técnicas para realizar la selección de las variables, entre las más importantes y más ampliamente usado es el método de regresión por pasos o stepwise.

#### **4.11.4 Método por pasos o stepwise**

Es un método automático que generalmente viene en los diferentes paquetes estadísticos como R, STATA y SAS.

Este método procede seleccionando o eliminando variables basado en algoritmos estadísticos que revisan la importancia de las variables en forma iterativa. La importancia de las variables se define en términos de significancia estadística de los coeficientes de los predictores.

En desenlaces binarios en donde se utiliza modelos logísticos, la selección por el método por pasos se utiliza una distribución binomial. En este caso, la importancia de las variables se evaluarían de acuerdo a la significancia estadística de los coeficientes de cada variable (66). En la regresión logística, los errores de la regresión siguen una distribución binomial y la significancia es evaluada a través de la prueba chi cuadrado. El procedimiento consiste en evaluar paso a paso las variables iniciando por las variables que muestren ser más importante comparando su efecto en el logaritmo de verosimilitud con el modelo que no contiene la variable (66).

Hay dos tipos de método de selección de stepwise: forward y backward. En la selección forward, se parte de un modelo sin variables y se va realizando la inclusión de las variables empezando por las más significativas de una a una. En cada paso, se evalúa la significancia estadística de la variable adicionada comparando su efecto en un modelo sin la variable evaluada. En la selección backward se parte de un modelo con todas las variables, pero contrario al método hacia delante (forward) se van retirando las variables del modelo de una a una de acuerdo a su significancia estadística (7,66).

La selección backward es la estrategia más usada porque presenta varias ventajas frente al forward. Primero, evalúa el efecto de la variable predictor teniendo en cuenta las otras variables al mismo tiempo. Segundo, las variables correlacionadas pueden mantenerse en el modelo, situación contraria en el método de forward (67). Sin embargo, otros autores opinan que al considerar en el modelo variables correlacionadas puede afectar el poder predictivo del modelo.

Por otro lado, una estrategia común en el desarrollo de modelos de predicción es realizar un análisis univariado antes del análisis multivariado. Las variables que cumplan con un valor  $p$  establecido por el investigador podrán ingresar al análisis final. Sin embargo, este método no es recomendado por algunos autores porque incrementa la selección de variables espurias, sobre-ajustamiento y modelos inestables (32).

Finalmente, un aspecto de gran importancia es tener en cuenta el conocimiento basado en revisiones de literatura combinado con la opinión de expertos en el tema.

#### 4.11.5 Manejo de valores perdidos

Los valores perdidos son definidos como valores no observados de una matriz de datos (68). Existen tres tipos de valores perdidos. Uno es cuando los valores perdidos ocurren en forma aleatoria (missing completely at random (MCAR)). Cuando la probabilidad de pérdida se debe a información no observada (missing not at random (MNAR)). El tercer grupo, es cuando la probabilidad de pérdida depende de información disponible (missing at random (MAR)). El caso más común en epidemiología son los valores perdidos tipo MAR debido a que hay variables que son no observables por otras variables que presenta el paciente (69).

En el primer caso, se realizan estrategias de imputación de tipo no condicional como por ejemplo la media. Para el segundo caso de MNAR no hay un método uniforme para manejar este tipo de valores perdidos. En el caso de los valores perdidos tipo MAR, por ser de tipo condicional se recomienda los métodos de imputación múltiple (69). Entre estos casos se dispone del método MICE. Este se basa en la densidad condicional de cada predictor dado con los otros predictores y el desenlace.

MICE traduce multiple imputation chained equations, en donde las imputaciones se realizan basado en una regresión donde la variable a imputar es la variable dependiente y las variables restantes (otros predictores y desenlaces) son las variables independientes. La regresión es utilizada para crear valores imputados en la variable con valores perdidos. Este método se repite varias veces, en el cual la variable con datos perdidos es imputada en sus diferentes repeticiones con valores distintos quedando finalmente una la imputación por promedio de los resultados obtenidos(70). El número de repeticiones para el proceso anteriormente descrito es definido bajo la regla de Rubin. La estimación de la imputación es el resultado de la computación del número de grupo de datos que se repite el proceso de imputación. Es decir, la estimación se calcula con la media de las  $n$  veces valores imputados. La varianza se calcula por (71) (Ecuación 2-12)):

$$T = W + \left(1 + \frac{1}{m}\right)B \quad (2-12)$$

Donde  $W$  es la varianza dentro los datos imputados,  $m$  es el número de veces que se repite la imputación y  $B$  es la varianza entre los datos imputados.

Muchos estudios realizan análisis de casos completos en donde se ignoran las variables con datos perdidos o con presencia de datos perdidos. Al desarrollar un modelo de predicción bajo estas condiciones afecta las estimaciones en el modelo desarrollado (72). Entre estos se puede mencionar:

- Sesgo en los coeficientes de regresión
- Puede afectar la capacidad discriminativa del modelo de predicción.
- Puede afectar en la estimación del factor etiológico.

La imputación múltiple puede generar sesgos pero en menor efecto que realizar un análisis de caso completo.

## 4.12 Validación

La validación busca verificar si el modelo de predicción trabaja adecuadamente en otros pacientes diferentes a quienes fueron usados para desarrollar el modelo. El desempeño de un modelo se realiza comparando la proporción de eventos observados con las estimadas (calibración) y la capacidad del modelo de diferenciar entre los pacientes con el evento entre los que no lo tienen (discriminación)(73).

Hay dos clases de validación: validación interna y validación externa. La validación interna mide la reproducibilidad del modelo predictivo en la misma población donde fue desarrollado. Es decir, la validación interna evalúa la validez para las condiciones donde los datos originales de los datos usados para el desarrollo del modelo. Por otro lado, la validación externa mide la generalizabilidad y transportabilidad del modelo de predicción en otras poblaciones plausiblemente relacionadas con la población de desarrollo del modelo. En este último, la población debe ser diferente a los datos de desarrollo y original pero relacionada en sus condiciones(7).

La validación interna presenta diferentes técnicas entre las que se puede numerar:

1. Validación interna aparente. La validación es realizada en el 100% de la muestra con la que fue desarrollado el modelo produciendo estimaciones optimistas, entendiéndose optimistas como la diferencia entre el desempeño obtenido en la muestra y el desempeño verdadero del modelo (7).

2. Validación Split-sample: Como su nombre lo indica, la muestra es dividida en forma aleatoria en dos partes. Una parte será utilizada para desarrollar el modelo y con la otra se realizará la validación interna. Generalmente, la parte para la validación esta entre 30% al 50% de la muestra. Este método de validación tiene varias desventajas como creación de modelos inestables, desbalance en la división de la muestra, encontrar pobre desempeño del modelo (7,74).

3. Validación cruzada: En la validación cruzada la muestra es dividida en deciles (1/10). El desarrollo del modelo se realiza con el 90% de la muestra. La validación es realizada en una de las décima parte de la muestra pero se repite varias veces con las otras partes para obtener resultados estables (7,74,75).

4. Validación con el método de bootstrap: Es un método de remuestreo que son tomadas de la muestra original con reemplazos en una parte de la muestra con elementos aleatorios. Por lo tanto, los nuevos grupos de datos tienen el mismo tamaño de la muestra original y están compuestas parte por datos muestra original y la otra parte por datos aleatorios. Este método calcula el optimismo comparando el desempeño del modelo en la muestra original y con las obtenidas de las otras muestras con reemplazos. Las ventajas que ofrece este tipo de validaciones son estimaciones menos sesgadas y con menor varianza (7,74,75).

En la validación externa, la población es diferente de la población de donde fue desarrollado el modelo de predicción. Se pueden encontrar varios tipos de validación (7):

1. Validación temporal: Esta validación consiste en que la población de validación se encuentra en diferentes puntos en el tiempo de la población utilizada para desarrollar el modelo de predicción a validar. Generalmente, los pacientes en la parte de validación son más recientes.

2. Validación geográfica: Las validaciones son realizadas con otras poblaciones provenientes de diferentes sitios geográficos que la población utilizada para desarrollar el modelo de predicción.

3. Validación completa independiente. Es la validación por investigadores independientes en donde pueden variar levemente las definiciones de predictores, desenlace y pacientes.

Los resultados de las validaciones indicarán como es el desempeño de los modelos y si los modelos requieren de re-calibración.

## 4.13 Tipos de medidas en el desempeño de modelos de predicción

### 4.13.1 Medidas de desempeño general

Desde el punto de vista estadístico, el desempeño general se mide a través de la distancia del evento observado con el evento estimado por el modelo de predicción. Para modelos de predicción logística, existen varias tipos de medidas. En este caso, se centrará en el  $R^2$  de Nagelkerke y el puntaje de Brier.

El  $R^2$  de Nagelkerke cuantifica la cantidad de información correcta en el modelo dado por el conjunto de datos. Para modelos no lineales y con el ajuste desarrollado por Nagelkerke, el  $R^2$  es calculado por(7):

$$R^2 = \frac{1 - (e^{\frac{-LR}{n}})}{1 - (e^{\frac{-2LL_0}{n}})} \quad (2-13)$$

Donde  $n$  es el tamaño de la población,  $LR$  es la relación de logaritmos de verosimilitud y  $LL_0$  es el log de verosimilitud del modelo sin variables. El  $LR$  sigue una distribución  $X^2$ , y puede ser calculado por la siguiente ecuación:

$$LR = -2(LL_0 - LL_1) \quad (2-14)$$

Donde  $LL_0$  corresponde a logaritmo de verosimilitud del modelo sin variables y  $LL_1$  es el logaritmo de verosimilitud del modelo con las variables predictoras (ver ecuación 2-11)

El  $R^2$  de Nagelkerke tiene la ventaja de estar en escala entre 0 y 100%. La relación entre la relación de logaritmos de verosimilitud y el  $R^2$  son aproximadamente lineal (7).

El puntaje de Brier es otra medida que calcula la diferencia en escala cuadrática de la proporción actual del desenlace y la proporción calculada. El puntaje va entre 0 (donde la predicción y la proporción de eventos observado es igual) a 1 (donde hay máxima diferencia). El puntaje de Brier es menos estricto que  $R^2$  de Nagelkerke en cuanto a penalizar las falsas predicciones cercanas a 0 o 100% (7,75,76).

$$Brier = Y * (1 - p)^2 + (1 - Y) * p^2 \quad (2-15)$$

Donde  $Y$  es la proporción actual de desenlace y  $p$  es proporción del desenlace calculada.

El puntaje de Brier es dependiente de la incidencia. Para una incidencia de 50% un puntaje de 0.25 no aporta información, en cambio para una incidencia de 10% un puntaje mayor a 0.09 se considera no útil (76).

#### 4.13.2 Calibración del modelo de predicción

La calibración mide el grado de acuerdo entre la proporción de desenlaces observados con la proporción de desenlaces estimados. Hay tres tipos de medidas: calibration-in-the-large, grafica de calibración y la prueba Hosmer-Lemeshow.

Con el método de grafica de calibración se realiza una comparación entre la predicción y la proporción de desenlaces observados en forma gráfica. Para desenlaces binarios, la muestra es dividida en cuantiles, en donde el eje x se coloca la media de la probabilidad estimada por el modelo de cada cuantil y en el eje y la media de la proporción observada del evento de cada cuantil. Idealmente, se espera obtener un ajuste de puntos que muestren una tendencia lineal con una pendiente de 45°. De acuerdo a esto, se espera un intercepto igual a 0 y una pendiente de 1 (7,76).

Otro medida de calibración es la pendiente de la calibración que es calculada a partir de una regresión logística, en donde, se establece la línea del predictor como una sola variable. La línea del predictor ( $lp$ ) está definido por el modelo de predicción (7,62).

$$\text{Logit}(P(y = 1)) = \alpha + \beta_i * x_i = lp \quad (2-16)$$

Entonces;

$$\text{logit}(P(y = 1)) = \alpha_{\text{calibración}} + \beta_{\text{calibración}} * lp \quad (2-17)$$

Donde  $\beta_{\text{calibración}}$  es la pendiente de calibración que idealmente debe ser igual a 1 y  $\alpha_{\text{calibración}}$  corresponde al intercepto que se espera sea igual a 0. Un valor de la pendiente por debajo de 1 indica sobreajustamiento, es decir, las estimaciones son bajas para predicciones bajas o son estimaciones altas para probabilidades altas. Si la pendiente es mayor que 1, significa que las probabilidades no son suficientemente altas. La principal causa de esto se debe al efecto de los predictores es diferente en la base de datos externa.

Cuando el intercepto es diferente de 0, esto indica que la predicción están sistemáticamente distantes de las probabilidades observadas (62,77).

La medida de calibration-in-the-large compara la media de la frecuencia estimada con media de la estimación de la probabilidad.

La prueba Hosmer Lemeshow o llamada también de bondad de ajuste, busca evaluar la habilidad de ajuste del modelo a los datos dados. Los pacientes usualmente son divididos por la probabilidad predictiva en deciles. La prueba de Chi-cuadrado es utilizada, en donde se compara el número de desenlaces esperados con el número de desenlaces observados. El problema con esta prueba es el pobre poder para detectar la falta de calibración en las validaciones externas o para detectar sobre ajustamientos de los predictores (7,66).

### 4.13.3 Discriminación

Otra propiedad a evaluar es la habilidad de diferenciar entre las personas con y sin el desenlace. La medida más común utilizada para medir la habilidad discriminativa es el estadístico C o estadística de concordancia. En desenlaces binarios el estadístico C es igual al área bajo la curva de ROC (receiver operating characteristic) (7,75,76,78).

La curva ROC es una gráfica que está compuesta en el eje de las X con la proporción de falsos positivo (1-especificidad) y en el eje Y con la proporción de verdaderos positivos (sensibilidad). El área bajo la curva representa la probabilidad que un paciente aleatoriamente seleccionado con el desenlace tenga más alta probabilidad que un paciente seleccionado independientemente sin el desenlace(78) . El rango del área está entre 0 a 1. De acuerdo a Hosmer Lemeshow, propuso que un área bajo la curva igual a 0.5 sugiere no discriminación, si está en 0.5 a 0.7 sugiere pobre discriminación, valores de 0.7-0.8 sugiere aceptable discriminación, valor entre 0.8 a 0.9 sugiere excelente discriminación (66).

Adicionalmente, como la curva ROC es una relación en alguna medida de la sensibilidad y especificidad, se puede mirar a que punto de corte el modelo presenta adecuada sensibilidad y especificidad (66).

## 4.14 Actualización

La actualización combina la información de un modelo de predicción previamente desarrollado y un nuevo grupo de pacientes. Este proceso es desarrollado cuando se encuentra en la validación externa modelos que no se ajustan a la nueva base de datos y es necesario la re-calibración a las nuevas características de los pacientes (33,79).

Hay diferentes métodos de calibración dependiendo del grado de actualización que se desee. Se tienen 7 métodos de actualización. En el método 1, se realiza la actualización del modelo de predicción a través del ajuste del intercepto. El método 2, se realiza actualización al intercepto y al coeficiente general del modelo de predicción. El método 1 y 2 se conoce como re-calibración del modelo, en el cual se intenta corregir la calibration-in-the-large. En el método 3, primero se realiza el método 2 y se evalúa la fuerza de cada predictor en la nueva base de datos para realizar el posterior. En el método 4, todos los

coeficientes de los predictores son reestimados. Éstos dos últimos métodos son denominados revisión del modelo. El método 5, consiste en realizar la re-calibración del modelo como lo realiza el método 3 pero se adicionan otras variables que se consideran importantes en la nueva base de datos. En el método 6, se realiza primero el método 4 y después se adicionan nuevos predictores según su significancia estadística. En el método 7, se realiza ajuste completo con las variables de las bases de datos y las originales del modelo. Los métodos 5 a 7 se consideran como un proceso de extensión del modelo (33,79)(Tabla 3).

**Tabla 3** Métodos de actualización de modelos de predicción con desenlace binario(7).

Método	Nombre	Notación
<i>Re-calibración</i>		
1	Actualización del intercepto	$\text{Logit}(p(Y = 1)) = \alpha + \text{offset}(lp)$ (2-18)
2	Re-calibración d intercepto y pendiente	$\text{Logit}(p(Y = 1)) = \alpha + \beta * lp$ (2-19)
<i>Revisión</i>		
3	Re-calibración + selectiva restimación	$\text{Logit}(p(Y = 1)) = \alpha + \beta * lp + \sum \gamma * x_i$ (2-20)
4	Restimación	$\text{Logit}(p(Y = 1)) = \alpha + \sum \delta * x_i$ (2-21)
<i>Extensión</i>		
5	Re-calibración + selectiva restimación + selectiva extensión	$\text{Logit}(p(Y = 1)) = \alpha + \beta + \sum \gamma * x_i + \sum \theta * z_i$ (2-22)
6	Restimación + extensión selectiva	$\text{Logit}(p(Y = 1)) = \alpha + \sum \delta * x_i + \sum \theta * z_i$ (2-23)
7	Restimación + extensión	$\text{Logit}(p(Y = 1)) = \alpha + \sum (\sigma * x_1 + \sigma * z_i)$ (2-24)

## 4.15 Meta análisis de modelos de predicción

Se han desarrollado dos métodos de agregación de modelos de predicción teniendo en cuenta que el meta-modelo estaría conformado por los diferentes modelos de predicción (34):

$$M = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} \text{Logit}^{-1}Lp1 \\ \text{Logit}^{-1}Lp2 \\ \text{Logit}^{-1}Lp3 \end{bmatrix} \quad (2-25)$$

Donde  $M_i$  indica cada uno de los modelos de predicción que integrarían el meta-modelo  $M$ .

#### 4.15.1 Método de modelo promediado

Este método se basa en establecer un promedio ponderado de cada modelo. El peso de cada modelo viene a depender del ajuste con la base de datos externa utilizada. Este método consiste en tres etapas (34):

Etapla primera. Los modelos son actualizados en la misma muestra externa para incrementar la comparabilidad entre los modelos. Se pueden realizar etapas de recalibración y revisión de los modelos.

Segunda etapa. Se calcula el promedio ponderado de cada modelo partiendo que el modelo promediado es calculado por la siguiente ecuación:

$$\bar{p} = \sum_{m=1}^m w_m * lp \quad (2-26)$$

Donde  $\bar{p}$  es el modelo promediado,  $w_m$  es el peso para cada modelo  $m_i$  y  $lp$  es la línea del predictor del modelo  $m_i$

Para calcular  $w_m$  se basa en el concepto de modelo promediado bayesiano, en donde los pesos son calculados a partir de las probabilidades posteriores de los modelos.

$$w_m = \frac{\exp(-0,5BIC_m)}{\sum_{l=1}^M \exp(-0,5BIC_l)} \quad (2-27)$$

Donde  $BIC_m = -2l_m + u_m \ln(N)$ ,  $u_m$  representa el número de parámetros de cada modelo determinado por las etapas en la actualización del modelo. Esta variable sería igual a 1 cuando solamente fue actualizado el intercepto, si se recalibro el modelo es igual a 2 por que se actualizó intercepto y pendiente. En caso de realizarse revisión del modelo, esta variable sería igual a  $K+1$  donde  $K$  es el número de variables que necesitaron ajustarse.  $l_m$  corresponde a la verosimilitud del modelo en la muestra de validación. El  $w_m$  debido a su forma de cálculo puede tener valores entre 0 a 1.

En la etapa 3, se seleccionan los modelos con  $w_m$  mayor a 0.0001. Se realiza un análisis de regresión lineal con variables independientes compuestos por los predictores de los modelos seleccionados y como variable dependiente la probabilidad promediada calculado por los modelos para el desenlace. Se utiliza la transformada logística para asegurar la linealidad de la variable dependiente.

El meta-modelo estará compuesto por:

$$M = \alpha + \sum_{k=1}^{K_{avg}} \beta_k x_{ik} + \epsilon \quad (2-28)$$

Donde M representa el meta-modelo, los valores de  $\alpha$  y  $\beta_k$  son calculados a partir de la estimación de máxima verosimilitud. Este modelo estará conformado por las variables provenientes de los modelos ponderados.

#### 4.15.2 Método de la regresión “stacked”

El segundo método es basado en las regresiones de “stacked” donde múltiples modelos son combinados. En este caso, no es necesario realizar la actualización de los modelos debido a que el método selecciona la mejor combinación en la muestra de validación (34).

En la regresión por “stacked”, considera cada modelo como un predictor del meta-modelo y se combinan en forma lineal.

$$M = \alpha_0 + \sum_{m=1}^m \alpha_m * Lp_{mi} \quad (2-29)$$

Donde  $\alpha_0$  y  $\alpha_m$  son los parámetros calculados por minimización de la función de error usando optimización ligada a restricciones. Estos parámetros tienen la restricción de ser mayores a 0. Si un modelo obtiene valores negativos ó 0, éste sería descartado. El único de parámetro que no estaría restringido es  $\alpha_0$ , para asegurar la línea basal de la muestra externa (34).

Finalmente, para el intercepto final se suman los interceptos ajustados de los modelos con el  $\alpha_0$  calculado por la regresión “stacked”. En los parámetros de las variables predictores se multiplica con  $\alpha_m$  estimado para cada modelo (34). Por ejemplo, para

el caso de intercepto final:

$$\beta_0 = \alpha_0 + \sum_{m=1}^m \alpha_m * \theta_m \quad (2-30)$$

En este caso, el coeficiente  $\beta_0$  es igual a la suma del parámetro  $\alpha_0$  y los interceptos de los modelos multiplicados por sus respectivos parámetros  $\alpha_m$ .

## 5. Metodología

Este trabajo fue desarrollado en dos etapas. Primero, se realizó una revisión sistemática de modelos de predicción para diagnóstico o tamizaje de cervicitis por *Chlamydia trachomatis*. Los modelos fueron evaluados en sus características de desempeño y riesgo de sesgo en el proceso de desarrollo y validación. Segundo, se seleccionaron modelos de predicción que cumplieran los criterios de selección previamente establecido. Posterior a esto, fueron agregados a través de dos métodos: modelo promediado y regresión “stacked”. En el primer método, cada modelo fue validado, recalibrado o revisado si este lo requería y finalmente se ejecutaba el modelo ponderado. En el segundo método, los modelos fueron agregados con los coeficientes originales. En ambos métodos se utilizó una base de datos externa proveniente de mujeres de Bogotá, Colombia.

### 5.1 Revisión sistemática de modelos de predicción clínica para el diagnóstico o tamizaje de cervicitis por *Chlamydia trachomatis*.

Se realizó una revisión sistemática de modelos de predicción para diagnóstico de infección cervical por *Chlamydia trachomatis*. La pregunta que se quería responder con la revisión sistemática fue planteada en términos PICOT (población, intervención, comparador, desenlace y tipo de estudio).

#### 5.1.1 Definición de la pregunta

A continuación se definen cada uno de los componentes de la pregunta PICOT.

##### Población

La población fue definida como mujeres mayores a 14 años de edad con diagnóstico de cervicitis por *Chlamydia trachomatis*.

##### Intervención

Las intervenciones fueron los modelos de predicción para tamizaje o diagnóstico de cervicitis por *Chlamydia trachomatis*.

## Comparador

Los comparadores eran cultivo, PCR, LCR que han sido definidos los patrones de oro para el diagnóstico de esta infección. Es importante aclarar, que se tomó en cuenta el cultivo debido a que antes de 1995 era considerado el patrón de oro.

## Desenlace.

Se hicieron dos tipos de evaluaciones en los estudios. Primero, se realizó una evaluación metodológica sobre el desarrollo y validación de los modelos de predicción. Segundo, se evaluó el desempeño de los modelos en términos de área bajo la curva ROC, sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, razón de verosimilitud positiva y negativa.

## Tipo de estudio

Solamente se consideraron estudios de tipo cohorte o corte transversal.

### 5.1.2 Búsqueda de métodos para la identificación de estudios

Las bases de datos electrónicas usadas para la localización de los estudios fueron Medline (Ovid), Embase y Bireme. La estrategia de búsqueda fue compuesta por términos MeSH, Emtree, Decs y términos relacionados con *Chlamydia trachomatis*, infección cervical e infección clamidial. Adicionalmente, se usó la estrategia de búsqueda para la identificación de estudios de predicción clínica que fueron sugeridos por Holland (80), Geersing (81) and Ingui (82).

### 5.1.3 Recolección de los datos

Dos investigadores seleccionaron los estudios para la inclusión en la revisión sistemática en forma independiente. Ellos revisaron el título y resúmenes de los artículos que fueron obtenidos en la búsqueda electrónica. En el caso de la decisión de inclusión no fuera posible a partir del resumen, se dirigía al artículo completo para tomar una decisión de inclusión o exclusión.

Los criterios de inclusión fueron establecidas en la pregunta PICOT. Los criterios de exclusión fueron mujeres trabajadores sexuales, mujeres embarazadas y mujeres con infección recurrente.

Los desacuerdos entre los investigadores fueron resueltos por consenso.

### 5.1.4 Evaluación de la metodología de la calidad

Se usó la herramienta PROBAST para la evaluación de la metodología de calidad del desarrollo de los modelos de predicción. Esta se encuentra compuesta por cinco dominios que son: Selección de participantes, desenlace, predictores, tamaño de la muestra y manejo de información y análisis de datos.

### **5.1.5 Evaluación del desempeño de los modelos de predicción**

Los modelos fueron comparados de acuerdo a las propiedades discriminativas o las características de desempeños que hubieran sido reportadas por el autor. Principalmente, se evaluó el área bajo la curva ROC, sensibilidad, especificidad, razón de verosimilitud positiva y negativa.

### **5.1.6 Criterio de selección de los modelos de predicción**

Los criterios de selección de los modelos de predicción para la agregación fueron:

1. Ser modelo de predicción
2. Usar la prueba de referencia para el diagnóstico de cervicitis por *Chlamydia trachomatis*
3. Incluir los coeficientes de cada predictor o disponer de los datos para poder calcular.
4. Contar con más del 50% de las variables dentro de la base de datos externa.

Los modelos que cumplían con estos criterios continuaban en las siguientes etapas.

## **5.2 Agregación de modelos de predicción para el diagnóstico de cervicitis por *Chlamydia trachomatis***

### **5.2.1 Método del modelo promediado**

Este método se realizó en tres etapas o fases. Primero, se actualizaron los modelos en la misma base de datos de pacientes para hacerlos comparables. Segundo, se calculó el peso para cada modelo y finalmente se desarrolló el meta-modelo (34).

#### **5.2.1.1 Validación y actualización de los modelos de predicción para el diagnóstico de infección cervical por *Chlamydia trachomatis*.**

##### **5.2.1.1.1 Validación de los modelos de predicción**

Se crearon las variables predictores de los modelos de predicción en la base de datos de las mujeres de Bogotá. La definición de cada variable corresponde exactamente a la forma como fue planteada en el artículo original. Respecto a los coeficientes de cada variable, se tuvieron que calcular cuando el artículo lo reportaba como un OR. Para éste cálculo se estimó el logaritmo de OR.

En la calibración se evalúa el acuerdo entre la proporción de desenlace observado y la probabilidad estimada. Para esto, se usó el método de grafica de calibración y la pendiente de calibración. Primero, se formó la línea del predictor de cada modelo.

Posteriormente, se estimó la probabilidad predictiva de cada paciente en la base de datos externa para cada modelo. Los pacientes fueron agrupados según los cuantiles de la probabilidad estimada.

En la gráfica de calibración, se graficaron las probabilidades observadas contra la probabilidad estimada. La media de la probabilidad estimada de cada cuantil fue comparada con la probabilidad observada en el mismo grupo (33). En esta gráfica, la relación de las medias en las probabilidades estimadas y calculadas debe tener una perfecta concordancia con la línea de referencia de 45°.

Por otro lado, se estimó el intercepto y la pendiente general de cada modelo usando el método de la pendiente de calibración (Ecuación 2-17).

$$\text{logit}(P(y = 1)) = \alpha_{\text{calibración}} + \beta_{\text{calibración}} * lp$$

De acuerdo a este método, un modelo se considera calibrado cuando el intercepto es igual a 0 y la pendiente es igual a 1. Si el modelo no estuviera calibrado, éste requeriría ser re-calibración.

### 5.2.1.1.2 Re-calibración de los modelos de predicción

Para la re-calibración del modelo, se usaron los resultados del método de la pendiente de calibración. Si el intercepto fue diferente de cero, entonces se adicionó el valor del  $\alpha_{\text{calibración}}$  al intercepto de la línea de predictor (79). Adicionalmente, cuando no hay calibración de la pendiente, es decir, la pendiente es diferente de uno, se re-calibró el modelo con la multiplicación de cada coeficiente original con la pendiente general estimada (33,79).

### 5.2.1.1.3 Revisión de los modelos de predicción

Debido a que puede haber efectos diferentes de algunas variables predictores en la base de datos externa comparado con la estimación del efecto en cada modelo, se ajustaría cada variable predictor en el modelo. De esta forma, se ajustarían las desviaciones que presenta cada modelo. Se usó el método de stepwise tipo forward, en el cual, se estableció una probabilidad de exclusión de 0.05 y la línea del predictor se establecía como forma "offset" (7). Por lo tanto, las covariables seleccionadas se corregirían con la desviación que se estima. De acuerdo a Steyerberg (33), la desviación en este caso, es la diferencia entre el coeficiente de predicción en la base de datos externa y el coeficiente de re-calibración. Por eso, la línea predictor actualizado sería (Ecuación 2-20):

$$\text{Logit}(p(Y = 1)) = \alpha + \beta * lp + \sum \gamma_i * x_i$$

Donde  $\alpha$  es el intercepto general,  $\beta$  pendiente general,  $l_p$  es la línea del predictor y  $\gamma$  es la desviación del predictor  $x$ .

#### 5.2.1.1.4 Evaluación de la medida de desempeño y habilidad discriminativa

Las medidas de desempeño representan la distancia entre la probabilidad estimada y la probabilidad observada (7). Se estimó el  $R^2$  que cuantifica la cantidad de variación de la base de datos externa explicada por el modelo predictivo. Adicionalmente, se calculó el puntaje de Brier que es estimado con la suma del cuadrado de las diferencias entre la probabilidad observada y la probabilidad estimada dividido por el tamaño de la muestra (7,33,76). Debido a que la prevalencia de pacientes con infección cervical en la base de datos externa es de 10 %, el apropiado puntaje de Brier debe ser por debajo de 0.09. Un puntaje de Brier igual a cero corresponde a un resultado ideal.

Para la habilidad discriminativa, se estimó el área bajo la curva ROC (76). Se consideró que un modelo con buenas propiedades discriminativas es cuando ésta es superior a 0.70.

#### 5.2.1.2 Estimación del peso de los modelos

De acuerdo al método antes descrito, los pesos de los modelos se calculan con la siguiente ecuación:

$$w_m = \frac{\exp(-0,5BIC_m)}{\sum_{i=1}^M \exp(-0,5BIC_i)}$$

Donde el BIC está dado por:

$$BIC_m = -2l_m + u_m \ln(N)$$

Donde  $l_m$  es log de la verosimilitud del modelo actualizado,  $u_m$  representa el número de parámetros estimados dependiendo del proceso de actualización.

Solamente, los modelos con peso mayor a 0.0001 fueron aceptados (34) .

#### 5.2.1.3 Estimación del meta-modelo

El meta-modelo fue construido con la predicción ponderada. Primero, se estimó la predicción ponderada. Posteriormente, se ejecutó un análisis de regresión lineal donde la variable dependiente era las predicciones ponderadas y las variables independientes eran las variables originales de cada modelo. Por lo tanto, el modelo estaría dado por (34):

$$y = \alpha + \sum \beta_i x_i + \epsilon_i$$

Donde la estimación de  $\alpha$  y  $\beta_i$  fue realizada por máxima verosimilitud.

## 5.2.2 Método de regresión “stacked”

### 5.2.2.1 Cálculo del peso de cada modelo

Al igual que el método de modelo promediado, se necesita estimar el peso de cada modelo, en este caso estos parámetros son denominados  $\alpha$ . Este fue estimado por minimización usando la optimización ligada a restricciones (34).

Los modelos se manejaron con sus coeficientes originales debido a que la regresión “stacked” simultáneamente actualiza y encuentra la mejor combinación de modelos en la base de datos externa (34).

Cuando el  $\alpha$  es igual a 0, el modelo no fue considerado.

### 5.2.2.2 Estimación del meta-modelo

Después de establecer que modelos que entran a formar el meta-modelo, el coeficiente de las variables predictoras es calculado a través de la multiplicación del parámetro  $\alpha$  del modelo con los correspondientes coeficientes originales de las variables predictoras que lo conforman. Para mejorar la capacidad de predicción y disminuir el riesgo de sobreajustamiento, los coeficientes son re-estimados con el método de penalización máxima de verosimilitud (83,84). En este método se calcula el factor de penalización que se aplicará a cada coeficiente. Este factor se estima con el método de criterio de información de Akaike (83).

### 3.2.2.3 Validación interna del meta-modelo

El meta-modelo fue validado por la técnica de bootstrap que permite evaluar el sesgo y ajustar la estimación de la exactitud predictiva con índices de corrección (4,75). Se condujo un análisis con 1000 replicaciones. El meta-modelo fue evaluado con los índices estadísticos: Puntaje de Brier,  $R^2$  generalizada, línea de calibración, intercepto y pendiente y la concordancia estadística.

#### **5.2.2.4 Imputación de variables predictivas con datos perdidos**

Los predictores con valores perdidos pueden influenciar en la selección de los modelos debido a que pueden producir sesgo en el cálculo de los coeficientes y el desempeño del modelo (72). Se utilizó el método de imputación con inferencia bayesiana usando 100 iteraciones (85). En cada imputación, la variable con datos perdidos fue tomada como variable dependiente y los predictores que pertenecían al modelo respectivo y al igual que el desenlace se consideraron como variables independientes.

#### **5.2.2.5 Software estadístico**

Se usó STATA versión 11 (86) y R Project versión 3.0.2 (87) con los paquetes Boot, Psych, Optimx, Mi y RMS.

### **5.3 Descripción de la muestra de validación**

Fue un estudio de corte transversal anidado de un estudio de cohorte de infección del tracto genital bajo femenino. La población de estudio fueron mujeres de tres centros en el año 2010 en la ciudad de Bogotá. Los criterios de inclusión fueron mujeres activas sexualmente entre 14-49 años, con cualquier síntoma de infección del tracto genital inferior y participación voluntaria en el estudio. Los criterios de exclusión fueron: condiciones médicas que requieren manejo a otro nivel de cuidado, actual estado de embarazo, historia de histerectomía, menstruación en el tiempo de examen ginecológico, previa participación del estudio, haber recibido terapia antibiótica legal o sistémica en los últimos 7 días y colocación previa de especulo vaginal y lubricante con aceite.

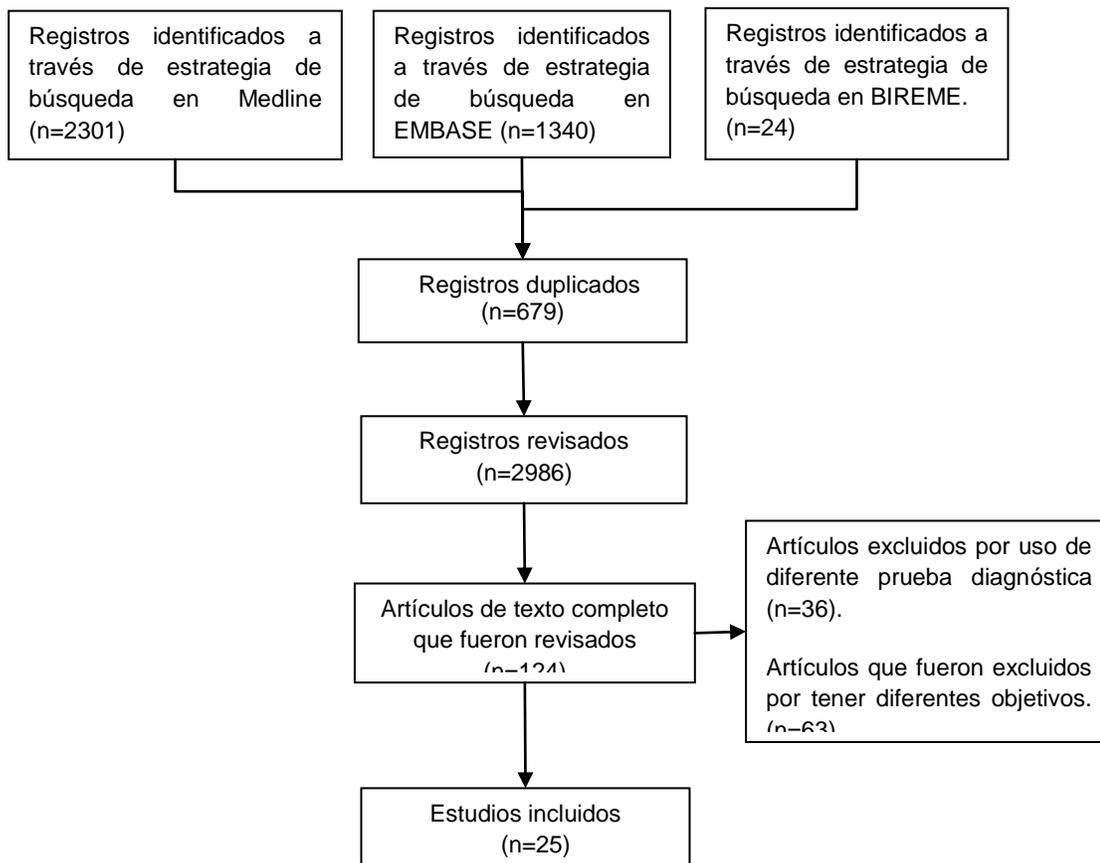
La muestra está compuesta por 1440 mujeres donde 134 mujeres fueron diagnósticas infección *Chlamydia trachomatis* con Polymerase Chain Reaction (PCR).



## 6.Resultados

### 6.1 Revisión sistemática de modelos predictivos para el diagnóstico o tamizaje de cervicitis por *Chlamydia trachomatis*.

La búsqueda realizada encontró 3665 referencias, en donde 679 de ellas eran duplicadas. 2961 estudios fueron excluidos por no cumplir con los criterios de inclusión y exclusión. Basado en el proceso de revisión del artículo completo en detalle, se identificaron 25 artículos que reunieron los criterios establecidos. (Figura 1)

**Figura 1** Diagrama de flujo prisma de la selección de estudios.

Dentro de los estudios excluidos, las razones de exclusión principalmente fueron: población diferente a la de población objetivo (exclusivamente trabajadoras sexuales, mujeres embarazadas, hombres, entre otros), ser otro tipo de estudios (principalmente estudios secundarios y de pruebas diagnósticas) y desenlaces relativos a otra patología. Se incluyeron 25 estudios que serán descritos en la siguiente sección.

### 6.1.1 Característica de los estudios

Los estudios incluidos fueron principalmente localizados en Norte América (12/25) y Europa (6/25). 10 de los estudios fueron ejecutados en centros diagnósticos de enfermedades de transmisión sexual y clínicas de planificación familiar debido a la facilidad de recolección de pacientes. 8 estudios usaron clínicas, hospitales, centros comunitarios u hospitales universitarios. Por otro lado, de los 25 estudios incluidos, 7 estudios incluyeron pacientes reclutados en prácticas de médicos generales o encuestas (Tabla 4).

Principalmente, la población consistió en pacientes que consultaban a atención médica para control de rutina ginecológica, cuidado en salud sexual o exámenes clínicos ginecológicos.

Los estudios con una población entre 16-50 años fue el más común de los estudios incluidos (19/25). Con adolescentes y adultos jóvenes (rango 16-30 años) fueron encontrados en 5 artículos. Solamente 1 estudio reportó estudiantes universitarias.

Los estudios con población adolescente, joven adulta y pacientes sintomáticos se caracterizaron por tener prevalencias superiores al 15%. La población general tuvo una prevalencia muy variable en donde 17 estudios tuvieron prevalencia por debajo del 10% de las cuales 11 estudios tuvieron prevalencia en un rango de 5-10%.

No se encontró ninguna relación entre la prevalencia y el sitio de atención de los pacientes. Los estudios con tamaños de muestra superiores a 200 pacientes tuvieron rangos de prevalencia de 2.8 – 10 % con una media de 7.46.

De acuerdo a los posibles objetivos que puede tener un estudio de modelos de predicción propuesto por el Dr Bouwmeester et al (32), 11 estudios tuvieron como objetivo evaluar predictores que fueran útiles para el diagnóstico o el tamizaje de la infección de la cervix por *Chlamydia trachomatis*. 6 estudios tuvieron el objetivo de desarrollar un modelo con validación interna y 8 estudios realizaron validación externa.

**Tabla 4** Localización y descripción de la población en los estudios incluidos de la revisión sistemática.

Características	Número de artículos (%)
<i>Localización</i>	
África	3 (12)
Asia and Oceanía	3 (12)
Europa	6 (24)
Norte América	12 (48)
Sur América	1 (4)
<i>Tipo de servicios</i>	
Población general	7 (28)
Centros diagnósticos de ITS y clínicas de planificación familiar	10 (40)
Centros comunitarios	2(8)
Clínica – hospital	4 (16)
Universidades	2(8)
<i>Población</i>	
Población general	19 (76)
Estudiantes universitarios	1 (4)
Mujeres adolescentes y adultos jóvenes	5 (20)

De los 25 estudios, 31 modelos fueron sugeridos para el diagnóstico y tamizaje de infección por *Chlamydia trachomatis*, de los cuales 15 fueron algoritmos y 16 fueron modelos logísticos (Tabla 5). 11 algoritmos presentaron validación externa mientras que 7 modelos logísticos tuvieron validación interna y validación externa. Los demás estudios no reportaron realizar algún tipo de validación.

**Tabla 5** Objetivos de los estudios incluidos y tipo de predicción clínica.

Característica de los estudios	Número de artículos
<i>Objetivos de los artículos</i>	
Predictor fue el objetivo primario	7 (28)
Predictor fue el objetivo secundario	4 (16)
Modelo de predicción con desarrollo y validación interna	6 (24)
Modelo de predicción con desarrollo y validación externa	1 (4)
Modelo predicción solo validación externa	7 (28)
<i>Tipo de predicción clínica</i>	
Algoritmo	15 (48.4%)
Modelo logístico	16 (51.6%)

Respecto al manejo de los predictores candidatos, la definición de éstos fue clara en todos los estudios. La mayoría de ellos fueron evaluados en forma ciega respecto al desenlace gracias a las características de las pruebas diagnósticas de clamidia las cuales necesitan un tiempo mayor de 24 horas para su interpretación.

La evaluación de la interacción entre los predictores no fue común en los estudios que utilizaron modelos logísticos. Solamente, en el 21,4% de los modelos fueron evaluadas las interacciones (Tabla 6).

La descripción de los datos perdidos fue encontrada en solo el 40% de los estudios. Principalmente, ellos reportaron el número de pacientes con datos perdidos en algunos de los predictores o del reporte de la prueba diagnóstica. De éstos, solo 6 artículos reportaron el manejo de datos perdidos usando el método de análisis completo. Este método no es recomendable por la pérdida de información y las asociaciones sesgadas producidas por el uso de datos altamente seleccionados (72,88). Ningún artículo reportó otro método de manejo de datos perdidos (Tabla 6). De acuerdo a recientes estudios, la imputación múltiple por ecuación en cadena es el método que ha mostrado mejores resultados que los otros métodos en cuanto a la estimación de los coeficientes y la habilidad discriminativa del modelo (70,72).

La variable continua más frecuente fue la edad, la cual fue transformada a una variable ordinal o dicotómica. La transformación ordinal fue la más común en los modelos logísticos. Sin embargo, si se analiza en conjunto con los algoritmos se encontró que la dicotomización pasa a primer lugar. Para esta transformación, en los estudios se seleccionaba un punto de corte con el objetivo de dicotomizar la variable, o se establecían diferentes rangos (Tabla 6).

**Tabla 6** Reporte de manejo de predictores en los estudios incluidos en la revisión sistemática

Características	Número de artículos
<i>Reporte y manejo de variables predictoras</i>	
Clara definición de predictores	25 (100)
Evaluación ciega de predictores	25 (100)
Evaluación de interacciones	3 (21.4*)
<i>Reporte de datos perdidos</i>	
No reporte	15(60)
Reporte	10 (40)
Reporte de método de manejo de datos perdidos data	6 (24**)
<i>Manejo de variables continuas*</i>	
Dicotomización	7 (39)
Categorización	11 (61)
Continua	0 (0)

\* Porcentaje calculada de estudios que desarrollaron modelos logísticos

\*\* Porcentaje calculada de estudios que reportaron datos perdidos

De acuerdo a la regla de 10 eventos por cada variable candidato, solo 8 artículos tienen problemas en el poder estadístico, que representa el 57,1% de los estudios con objetivo de desarrollar un modelo de predicción (tabla 7). Los estudios de Harrison (29), Sellors (89) y Verhoeven (31) tuvieron una relación evento-predictor debajo de 5, es decir, por cada predictor hubo 5 eventos. Los estudios de Al-Tayyib (90), Van Valkengoed (91), Stein (92), Duynhoeven (93), Rudpao (28) y Johnson (94) tuvieron una relación por encima de 10, por lo tanto, tiene mayor poder estadístico.

Por otro lado, referente al tamaño de la muestra en los estudios focalizados en realizar validación externa, el estudio de Valkengoed (91) reclutó 58 pacientes con evento. De acuerdo a Vergouwe (95) este número de eventos es insuficiente para ésta clase de estudios.

**Tabla 7** Tamaño de la muestra de los estudios incluidos en la revisión sistemática

Relación desenlace - número de	Número de artículos (%)*
<5	3 (21.4)
5-10	5 (35.7)
10-15	3 (21.4)
15-20	3 (21.4)

\*Porcentaje calculada de los estudios incluidos con el objetivo de desarrollar un modelo logístico.

Se encontró que todos los estudios incluidos que ejecutaron selección de predictores usaron análisis univariable como primer filtro. En este análisis, el 50% de los artículos usaron valores p inferiores a 0.05. Otros 4 estudios, no reportaron el punto de corte para la inclusión de variables (Tabla 8).

Para la selección de variables, el método de stepwise fue usado en 15 (83,3%) de 18 artículos que reportaron este análisis. Los estudios con este método, el stepwise tipo backward fue el más frecuente usado comparado con el tipo forward. Cuatro artículos no reportaron el tipo de método de stepwise usado (Tabla 8). El valor p más frecuente fue valores inferiores a 0.05. El estudio de Van Duynhoeven(93) no reportó método de selección multivariable.

**Tabla 8** Criterio y método de selección de variables de los artículos incluidos en la revisión sistemática.

Características	Número de artículos (%)
<i>Selección de predictores en modelos multivariados</i>	
Análisis univariado	18 (100)
<i>Criterio de selección en análisis univariado</i>	
Valor p <0.05	9 (50)
Valor p > 0.05	5 (27.8)
No reporta	4 (22.2)
<i>Selección multivariable</i>	
Stepwise tipo forward	3 (16.6)
Stepwise tipo backward	8 (44.4)
Stepwise sin especificar tipo	4 (22.2)
Método no descrito	1 (5.5)
No realiza análisis multivariado	2 (11.1)
<i>Criterio de selección de análisis multivariado</i>	
Valor p <0.05	8 (53.3)
Valor p > 0.05	1 (6.6)
No reporta	6 (40)

Trece reglas de predicción clínica tuvieron validación externa entre modelos logísticos y algoritmos, sin embargo los algoritmos fueron más frecuentes. La descripción de la habilidad discriminativa y del desempeño de los modelos y algoritmos fue incompleta ya que se limitaron a medir sensibilidad y especificidad.

Para los modelos de predicción solo se midió la habilidad discriminativa con el área bajo la curva ROC. Las medidas de calibración y de desempeño general no fueron reportadas en la gran mayoría de los artículos tanto con validación externa como validación interna. Solamente, Harrison (29), Verhoeven (31) y Shields (96) reportaron el valor de la estadística Chi-cuadrado de la prueba de bondad de ajuste. El 42.8% de los modelos desarrollados tuvieron validación interna donde el método más utilizado fue el de Split simple (Tabla 9).

**Tabla 9** Descripción de la validación de los reglas de predicción en los estudios incluidos en la revisión sistemática.

Validación de las reglas de predicción	Número de CPR (%)
Validación externa	13 (41.9)
Algoritmos	11
Model	2
Validación interna	6 (42,8*)
Boostrapling	2
Split sample	4

\*Porcentaje de estudios incluidos con objetivo de desarrollo de modelo logístico.

Desde una perspectiva más global, se pueden observar en la tabla 10 las variables más frecuentes que fueron incluidas en el modelo. La edad fue la más reportada, la cual de 31 reglas de predicción, 21 utilizaron esta variable predictora. El punto de corte más frecuente en ésta fue 25 años.

Otras variables más frecuentemente usadas fueron: más de un compañero sexual, nuevos compañeros sexuales, no estar casado, tener un compañero sexual con síntomas. En las variables clínicas sobresale friabilidad cervical, mucopus, flujo vaginal, número de leucocitos, entre otros. Adicionalmente, se encontraron factores predictivos relacionados con el uso de anticonceptivos, especialmente en forma oral y condón. Sin embargo, esta frecuencia de uso de las variables predictores no indica que unas sean más fuertes que las otras variables.

**Tabla 10** Variables de predicción frecuentes en las reglas de predicción incluidas.

Variable Predictores	Numero (%)
Edad	21 (68)
Más de 1 compañero sexual	14 (45)
Más de 1 compañero sexual en los últimos 12 meses	7
Más de 1 compañero sexual en los últimos 3 meses	5
Más de 1 compañero sexual en los últimos 6 meses	2
Nuevo compañero sexual	12 (39)
Nuevo compañero sexual en los últimos 3 meses	8
Nuevo compañero sexual en los últimos 2 meses	3
Nuevo compañero sexual en los últimos 12 meses	1
No casado	10 (32)
Compañero con síntomas	11 (35)
Friabilidad cervical	13 (42)
Mucopus	8 (26)
Flujo Vaginal	6 (19)
Numero de leucocitos	5 (16)
Flujo cervical	4 (13)
Ectopia cervical	4 (13)
Eritema cervical	4 (13)
Disuria	3 (10)
Relacionado con anticoncepción	5 (16)
Uso de Condón	4 (13)

### 6.1.2 Riesgo de sesgo en los estudios incluidos

Se evaluaron los riesgos de sesgo en los siguientes dominios (Figura 2).

#### *Selección de participantes*

Los estudios de Harrison(29), Wang (97), Verhoeven (98) y Passey (99) presentaron bajo riesgo de sesgo de selección en este dominio. Otros estudios tuvieron insuficiente información referente a los criterios de inclusión y exclusión. Por otro lado, el principal tipo de recolección de pacientes había sido no aleatorio o no consecutivo. Respecto a la aplicabilidad, algunos modelos de predicción tuvieron alto riesgo de no ser considerados debido a que recolectaron pacientes de alto riesgo como adolescentes y mujeres sintomáticas.

#### *Desenlace*

Los estudios usaron PCR, LCR o cultivo como prueba diagnóstica de referencia para infección clamidial que fue un criterio establecido a priori. Este grupo de pruebas son o fueron considerados el patrón de oro en el diagnóstico de esta infección (41,48,49). Por ende, muchos de los estudios presentaron bajo riesgo de sesgo en este dominio. Sin embargo, 5 de los artículos incluidos tuvieron combinación de diferentes pruebas diagnósticas que pueden producir sesgo en la detección del agente infeccioso.

#### *Predictores*

Se evaluó la medida y el momento de las variables predictoras. El 65 % de los estudios ejecutaron esta medida antes del diagnóstico de la infección por clamidia y fueron ciegos a otra información. Además, ellos definieron a priori los predictores a partir de encuestas estructuradas y semi-estructuradas. El 35% de los artículos no reportaron suficiente información para evaluar este dominio. Principalmente, la falta de información fue relacionada al momento de medición de las variables y la persona quien realizaba la encuesta.

#### *Tamaño de la muestra y flujo*

De los 25 estudios incluidos, 10 estudios tuvieron tamaño de muestra de al menos 10 eventos por cada candidato predictor. Adicionalmente, muchos de los estudios no reportaron datos perdidos y manejo de estos. Respecto a la medición de desenlace y los predictores, hubo bajo riesgo por el tipo de diagnóstico que se ésta manejando. Ninguno de los estudios reportó proporcionar alguna intervención a los participantes durante la medición de las variables predictoras. Sin embargo, por los tipos de variables predictoras no sería tan relevante este tipo intervención.

#### *Análisis de datos*

18 estudios tuvieron alto riesgo de sesgo en el análisis y desarrollo de modelos de predicción. La principal falla corresponde a estudios que no responden al no reporte de validación interna, estimación penalizada de coeficientes y medidas de calibración in la validación externa. Por otro lado, muchos de los estudios realizaron estrategias de selección de análisis multivariable. Dos estudios (La montagne(100) y Van Valkengoed (101)) tuvieron modelos que no eran consistentes con el análisis multivariable. El estudio de Passey (99) no fue claro en el proceso de desarrollo del modelo.

Respecto a la aplicabilidad, 12 estudios se categorizaron como no claros debido a la falta de información sobre sobreoptimismo y medidas de desempeño de los modelos causados por la falta de validación de los mismos. Dos estudios tuvieron alto riesgo de no aplicabilidad por que presentaban alguna diferencia entre el modelo final y el análisis multivariable.

**Figura 2** Evaluación de la calidad metodológica de los estudios incluidos. (El significado de los colores son: Verde: bajo riesgo de sesgo; amarillo: no es claro; rojo: alto riesgo de sesgo)

Autor	Año	Dominio 1. Selección participantes		Dominio 2. Desenlace		Dominio 3. Predictores		Dominio 4. Tamaño de la muestra y flujo		Dominio 5. Análisis de datos	
		Riesgo de sesgo	Aplicabilidad	Riesgo de sesgo	Aplicabilidad	Riesgo de sesgo	Aplicabilidad	Riesgo de sesgo	Aplicabilidad	Riesgo de sesgo	Aplicabilidad
Alary (102)	1998	Verde	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Altayyid (90)	2008	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Fonck (103)	2000	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Guimarañes (104)	2009	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Johnson (94)	1990	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Latino (105)	2008	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
La montagne (100)	2004	Rojo	Verde	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Masse (106)	1991	Rojo	Verde	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Miller (107)	2000	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Passey (99)	1998	Verde	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Reed (108)	2007	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Rugpao (28)	2010	Verde	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Ryan (109)	1998	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Sellors (89)	1992	Verde	Verde	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde
Van Valkengoed (101)	2000	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Verhoeven (31)	2003	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Verhoeven (110)	2004	Verde	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Van Duynhoven (93)	1995	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Stein (92)	2008	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Wang (97)	2003	Verde	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Hocking (111)	2005	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Stergachis (30)	1993	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Valkengoed (91)	2000	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Shields (96)	2004	Rojo	Rojo	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde
Harrison (29)	1985	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde	Verde

La figura 2 demuestra que los modelo que las reglas de predicción presentan alto riesgo de sesgo en los dominios de selección de los participantes, tamaño de la muestra y

análisis de datos. Los otros dos dominios presentan bajo riesgo de sesgo por el comportamiento del proceso diagnóstico de la infección.

### 6.1.3 Desempeño de los modelos

En la tabla 11 se reporta el desempeño de las reglas de predicción clínica en los estudios incluidos. Principalmente, los reportes fueron relativos a la sensibilidad y especificidad sobresaliendo la falta de reporte de las características operativas.

**Tabla 11** Desempeño de las reglas de predicción clínica en estudios incluidos en la revisión sistemática

Autor	Model	Sensi.	Especi.	VPP	VPN	AUC	LR+	LR-	Tipo de regla de predicción
Alari	-	86.7	41.8	11.2	97.4	-	1.49	0.32	Algoritmo
Fonck	-	66	54	24	-	-	1.43	0.63	Algoritmo
La Montagne	-	93.3	-	-	-	-	ND	ND	Algoritmo
Miller	CDC	85	38	-	-	-	1.37	0.39	Algoritmo
	Seattle-1	56	34	-	-	0.599	0.85	1.29	Algoritmo
	Wisconsin	50	66	-	-	0.604	1.47	0.76	Algoritmo
	Ontario	76	42	-	-	0.63	1.31	0.57	Algoritmo
	California-1	94	20	-	-	0.633	1.18	0.30	Algoritmo
	California-2	97	9	-	-	0.701	1.07	0.33	Algoritmo
Passey		81	63	45	-	-	2.19	0.30	Algoritmo
Reed		75	71	56	85	-	2.59	0.35	Algoritmo
Ryan		91	35.6	32.6	-	-	1.41	0.25	Algoritmo
Sellers		87.5	-	-	-	-	ND	ND	Algoritmo
Verhoeven	Varhoeven	-	-	-	-	0.7	ND	ND	Algoritmo
Wang	-	90.8	46.9	50.9	89.3	-	1.71	0.20	Algoritmo
Al-tayyid	-	60	68	-	-	0.688	1.88	0.59	Modelo
Guimarañes	WHO (Score risk)	31.9	76.5	20.8	85.4	-	1.36	0.89	Modelo
Harrison	-	91	95	65	99	-	18.2	0.09	Modelo
Hocking	-	86	45	-	-	-	1.56	0.31	Modelo
Johnson	-	-	-	-	-	0.698	ND	ND	Modelo
Latino	-	-	-	-	-	0.732	ND	ND	Modelo
Masse	-	84	-	-	-	-	ND	ND	Modelo

Autor	Model	Sensi.	Especi.	VPP	VPN	AUC	LR+	LR-	Tipo de regla de predicción
Miller	Seattle-2	84	51	-	-	0.726	1.71	0.31	Modelo
	Seattle-3	92	31	-	-	0.723	1.33	0.26	Modelo
Rugpao	-	72.8	59.8	3.4	99.1	-	1.81	0.45	Modelo
Shields	-	-	-	-	-	-	ND	ND	Modelo
Stein	-	76.6	52.4	-	-	0.7	1.61	0.45	Modelo
Stergachis	-	88	56	6	-	-	2.00	0.21	Modelo
Valkengoed	Stergachis	80.8	-	-	-	0.67	ND	ND	Modelo
	Gaydos	73.7	-	-	-	0.6	ND	ND	Modelo
Van Duynhoven	-	79	-	35	-	-	ND	ND	Modelo
Van Valkengoed	-	63	-	-	-	0.67	ND	ND	Modelo
Verhoeven	-	92.3	37.5	-	-	0.88	1.48	0.21	Modelo

Razón de verosimilitud calculada por los autores

Hay un amplio rango de los valores de sensibilidad y especificidad entre los modelos y algoritmos. El comportamiento más característico es tener sensibilidad alta y especificidad baja. Por lo tanto, las razones de verosimilitud entre los estudios son bajas. El modelo de Harrison tuvo el más alto valor de la razón verosimilitud positivo y el más bajo razón de verosimilitud negativo. Los otros modelos tuvieron una razón de verosimilitud positiva inferior a 2.5.

Sin embargo, estos modelos y algoritmos fueron desarrollados para tamizaje de infección clamidial, por eso, los investigadores buscaban mayor sensibilidad que especificidad. Por otro lado, la habilidad discriminativa que fue medida con el área bajo la curva ROC, fue baja para todos los modelos. Solamente el modelo de Verhoeven reportó un área bajo la curva de 0.88, la cual es considerada como buena exactitud diagnóstica.

La mayoría de los modelos y algoritmos no reportaron los valores predictivos positivos ni negativos. Para los modelos que reportaron éstos parámetros mostraron precisiones bajas.

#### 6.1.4 Selección de modelos para la agregación

De los 16 modelos de predicción clínica, 7 de ellos reunieron los criterios de inclusión. Los algoritmos no fueron incluidos debido a que no tenían coeficientes de cada variable. La razón de exclusión de los otros modelos de predicción clínica fueron: tres excluidos (Latino (105), Masse (106) y Sellors (89)) por diagnosticar la infección cervical por

*Chlamydia trachomatis* con más de una prueba diagnóstica diferentes a LCR, PCR y cultivo y seis modelos de predicción fueron excluidos (Guimarañes (104), Johnson (94), Rudpao (28), Van Valkengoed (101), Shields (96) y Valkengoed (91)) debido a que tuvieron menos del 50% de las variables en la muestra de validación.

### 6.1.5 Descripción de los modelos incluidos

En la tabla 12, se describen las características de los modelos seleccionados. En esta tabla se encuentra año de publicación, descripción de la población, localización del estudio, tamaño de la muestra, sitio de recolección de pacientes, y prevalencia de la infección cervical por *Chlamydia*.

Seis estudios tuvieron prevalencias en un rango de 3.7 a 8 mientras que el estudio de Duynhoven tuvo una prevalencia de 21.5. Respecto al sitio de recolección de pacientes, 5 estudios fueron ejecutados en clínicas y hospitales. Los otros estudios fueron encuestas a la población general y a consultorios de práctica general. El rango de edad fue de 18-50 años. Hubo un estudio con mujeres adolescentes, dos estudios con mujeres adultas y tres estudios en población general. Adicionalmente, el porcentaje de mujeres solteras fue similar entre los estudios. El rango de mujeres con síntomas fue entre 28-50%.

**Tabla 12** Características de los modelos incluidos en los procesos de agregación.

Autor	Año	Población	Localización	Sitio de recolección	Tamaño de la muestra	Prevalencia
Al-Tayyib	2008	Mujeres entre edades de 18-35 años con media de 23.2 años y desviación estándar de 4.6; EL 70% de las mujeres eran solteras. Todas las mujeres eran sexualmente activas en los últimos 90 días. El 40% de los pacientes tuvieron síntomas.	Baltimore	Departamento de emergencia	3008	7.5
Harrison	1985	Mujeres estaban entre las edades 17-45 años y con una media 23.3 años. Ellos fueron principalmente estudiantes universitarias. El 79% de las mujeres eran solteras. Ellas habían asistido a consulta por rutina, síntomas o anticoncepción.	Arizona	Clínica ginecológica	160	8

Autor	Año	Población	Localización	Sitio de recolección	Tamaño de la muestra	Prevalencia
Hocking	2005	Mujeres estaban entre un rango de 15-40 años. La información fue recolectada en clínicas EL 40% de las mujeres había asistido por síntomas relacionadas a ITSs.	Australia	Clínicas de Infección de transmisión sexual	1782	4
Stergachis	1994	Mujeres estaban entre rango de edad entre 15-34 años, con una media de 28 años. El 40% de las mujeres eran solteras. Adicionalmente, el 48% de ellas tenían síntomas.	Seattle	Hospital	1804	3.7
Stein	2008	Mujeres entre la edad 18-26 años con media de 21.4 años. La información fue recolectado por encuestas. El 28,7% tuvieron síntomas. El 76,6% eran mujeres solteras.	Carolina Norte	Población general	5252	5.1
Verhoeven	2003	Mujeres entre 14-40 años con media de 25,2 años. EL 31.7% de las mujeres eran solteras. Eran elegibles para participar si ellas eran menores de 40 años y habían tenido actividad sexual el último año.	Belgica	Población general	787	5
Van Duynhoven	1995	Las mujeres eran entre 15-53 años con una media de 28 años. 14.8 % tuvieron historia de enfermedades de transmisión sexual; el 50% de las pacientes tenían diagnóstico actual de infecciones de transmisión sexual.	Amsterdan	Clínicas de Infección de transmisión sexual	684	21.5

### **6.1.6 Descripción de las variables de los modelos incluidos**

En la tabla 13 se enlistan cada modelo con sus respectivas variables y coeficientes.

Se puede notar en los modelos incluidos que sus factores predictivos están compuestos por características demográficas, comportamientos sexuales de riesgo y signos y síntomas, a excepción de los modelos de Hocking y Stein que no incluye síntomas.

Los modelos seleccionados presentan variables que son comunes entre los modelos como la edad, número de leucocitos, número de compañeros sexuales, uso de anticonceptivos, compañero sintomático, disuria y friabilidad cervical. Sin embargo, tienen diferentes puntos de corte o criterios de clasificación entre cada variable. Por ejemplo, la edad, se encuentra en 6 modelos de los 7 incluidos, pero en los modelos de Stein y Varhoeven se encuentra como una variable categórica y en los otros como variable dicótoma. Un caso similar ocurre con el número de compañeros sexuales en el cual en el modelo de Hocking es usado como variable dicotómica mientras que en los modelos de Stein y Varhoeven esta como variable categórica.

**Tabla 13** Descripción de las variables de cada modelo seleccionado para el proceso de agregación

Autor	Año	Variables	Coefficiente	SE	OR	IC 95%
Al-Tayyib	2008	Edad <25 años	-	-	2.94	2.12 - 4.09
		Secundaria o menor	-	-	2.05	1.35 - 3.12
		Nuevo compañero sexual en el último año	-	-	1.97	1.41 - 2.77
		Disuria	-	-	1.38	1.04 - 1.83
Harrison	1985	Puntaje de cervicitis	1.149	0.359	-	-
		Contracepción oral	2.635	0.814	-	-
		Compañero con uretritis	1.861	0.623	-	-
		Ninguna contracepción versus oral	-0.608	0.685	-	-
		Otra anticoncepción versus ninguna	-1.466	0.729	-	-
		Constante	-4.078	1.685	-	-
Hocking	2005	Compañero con infección	-	-	10.7	5.4 - 21.3
		Numero de compañeros en el último año	-	-	2.5	1.4- 4.7
		Condomes < 50% de las veces	-	-	2.5	1.3 -4.7
		Edad <= 25	-	-	1.5	0.9 - 2.6
		Uso de drogas inyectables (<1 año)	-	-	3.1	0.8 - 11.9
Stergachis	1994	Soltera	-	-	5	1.5 - 17.2
		Ectopia cervical	-	-	3.7	2 - 6.9
		Duchas en el último 12 años	-	-	2.2	1.1 - 4.2
		Nuliparas	-	-	2.1	1 - 4.3
		Edad>=24	-	-	2	1.1 - 3.8
		>=2 compañeros sexuales en el último año	-	-	1.9	1 - 3.7
		Afroamericano	-	-	3.1	1.4 - 6.9
Stein	2008	Numero de compañeros en el último año (1)	1.4	-	4	1.5- 10.9
		Numero de compañeros en el último año (2-50)	2.01	-	7.4	2.7 - 20.5
		Percepción de riesgo de ITS (Moderado o alto)	1.3	-	3.7	2.1 - 6.3
		No ser estudiante	0.68	-	2	1.3 - 3
		Edad (18-21)	0.43	-	1.5	1.1 - 2.2

Autor	Año	Variables	Coficiente	SE	OR	IC 95%
		Edad (25-26)	0.74	-	2.1	1.2- 3.7
		Soltero	1.03	-	2.8	1.5 - 5.2
		Nunca embarazada	0.57	-	1.8	1.2 - 2.6
		No anticoncepción	0.44	-	1.5	1 - 2.44
		Constante	-6.64	-	-	-
Varhoeven	2003	Edad14-17	-0.634	1.184	-	-
		Edad 18-22	1.422	0.537	-	-
		Edad 23-27	1.284	0.526	-	-
		Numero de compañeros en el último año (2)	1.79	0.499	-	-
		Numero de compañeros en el último año (3-5)	2.178	0.552	-	-
		Numero de compañeros en el último año (>=6)	2.27	0.642	-	-
		Ninguna anticoncepción	2.438	0.441	-	-
		Disuria	0.883	0.429	-	-
		Compañero con alguna síntoma urinario	2.05	0.534	-	-
		Sangrado poscoital (Frecuente)	3.048	0.984	-	-
		Constante	-1.979	7.355	-	-
Van Duynhoven	1995	Edad < 26	-	-	1.73	1.05- 2.85
		Historia de enfermedad de transmisión sexual	-	-	0.56	0.35 -0.9
		Friabilidad cervical	-	-	2	1.19 - 1-34
		Mucopus	-	-	1.71	1.07 -2.73
		Células guía	-	-	2.26	1.44 - 3-55
		Numero de leucocitos >= 10	-	-	4.09	2.31 - 7.22

## 6.2 Agregación de modelos de predicción para el diagnóstico de cervicitis por *Chlamydia trachomatis*

### 6.2.1 Descripción de la muestra externa de validación

La muestra externa usada para la validación, actualización y meta-análisis de los modelos tiene un tamaño de muestra 1381 pacientes con 133 mujeres diagnosticadas con infección clamidial. Los datos sociodemográficos indican que la edad tuvo una distribución normal con una media de 31.5 y una desviación estándar de 8.92. El 43.67% de las mujeres fueron solteras. Un 73,64 % de las mujeres no usan algún método de anticoncepción (Tabla 14).

**Tabla 14** Características de la muestra externa en las características demográficas, comportamiento sexual y síntomas.

Características	Muestra (n=1381)	Muestra sin trabajadoras sexuales (n=1266)	Trabajadores sexuales (n=114)	Pacientes infectadas (n=133)
Número de eventos	133 (9.63)	117 (9.24)	16 (14)	
<i>Características demográficas</i>				
Edad	31.35 (8.92)	31,62 (8,99)	28,34 (7.55)	27.14 (7.46)
Educación (Inferior a secundaria)	581 (42.07)	553 (43.68)	86 (75.44)	75 (56.39)
<i>Estado civil</i>				
Soltero	593 (43.67)	521 (41.15)	72 (63.16)	71 (53.79)
Casado/Cohabitante	613 (45.14)	606 (47.87)	31 (27.20)	51 (38.64)
Otro	152 (11.20)	139 (10.98)	11 (9.65)	10 (7.58)
<i>Comportamiento sexual</i>				
No uso de anticoncepción	1017 (73.64)	933 (73.7)	83 (72.80)	93 (70.45)
Uso de condón				
Nunca	694 (64.8)	650 (64.9)	44 (62.86)	63 (60.58)
Algunas veces	282 (26.33)	260 (26)	22 (31,43)	34 (32.69)
Siempre	95 (8.87)	91 (9.1)	4 (5.71)	7 (6.73)
Compañero estable	1074 (77.76)	1004 (79.30)	70 (52.6)	104 (78.19)
Nuevo compañero sexual últimos 12 meses	48.46 (271)	1.26 (1.60)	577 (72.4)	31,4 (137)
<i>Síntomas y antecedentes</i>				
Historia de STD	428 (31.15)	395 (31.3)	33 (29,2)	43 (32.8)
Flujo vaginal	1,242 (89.93)	1145 (90.44)	96(84.2)	127 (95.48)
Mucopus	120 (8.70)	95 (7.51)	25 (22.12)	20 (15.04)
Friabilidad	137 (9.93)	116 (9.17)	21 (18.58)	20 (15.04)

En la muestra general, hubo 8,24% de trabajadoras sexuales. Ellas tuvieron diferentes características que la población general. Este grupo presentó un porcentaje mayor de cervicitis por clamidia; hubo un alto porcentaje de mujeres que tuvo educación inferior a estudios de secundaria; el porcentaje de mujeres casadas fue inferior en comparación al grupo de mujeres no trabajadoras sexuales; el porcentaje de mujeres solteras fue más alta; los nuevos compañeros sexuales en los últimos 12 años tuvieron mayor proporción; y estas mujeres revelan un alto porcentaje de mucopus y friabilidad al compararse con las mujeres no trabajadoras sexuales.

Sin embargo, el tamaño de la muestra de las mujeres trabajadoras sexuales es muy baja como para comparar y establecer afirmaciones acerca de este grupo de población.

### **6.2.2 Validación y actualización de los modelos seleccionados**

Los modelos seleccionados fueron validados, re-calibrados y actualizados. Los modelos de Al-Tayyib (90) y Van Duynhoeven (93) no fueron revisados debido a que no hubo mejoramiento del Log de verosimilitud (Log-likelihood). La tabla 15 muestra el cambio en los valores de coeficientes durante las diferentes etapas.

Los modelos de Harrison, Hocking, Stein, Stergachis y Varhoeven requirieron el proceso de revisión. El modelo que requirió mayor número de estimación de coeficientes fue Varhoeven con 5 variables. Las variables más frecuentes que requirieron estimación de la última etapa de actualización de los modelos incluidos fueron la edad, número de compañeros sexuales, uso de anticoncepción y nulípara.

**Tabla 15** Cambio de valores de coeficientes durante validación y actualización de los modelos incluidos.

Modelos	VARIABLES predictoras*	Coefficientes iniciales	Coefficientes re-calibración	Coefficientes revisión**
Altayyib 2008	Log-likelihood	-429.355	-426.181	
	Edad	1.078	0.671	
	Secundaria o menor	0.718	0.447	
	Nuevo compañero sexual en el último año	0.678	0.422	
	Disuria	0.322	0.200	
	Intercepto		-3.192	
	Pendiente calibración		0.622	
Harrison 1985	Log-likelihood	-532.801	-424.728	-418.138
	Puntaje de cervicitis	1.149	0.261	<b>0.437</b>
	Contracepción oral	2.635	0.598	0.598
	Compañero con uretritis	1.861	0.423	0.423
	<b>Ninguna contracepción versus oral</b>	-0.608		
	Otra anticoncepción versus ninguna	-1.466	-0.333	<b>0.301</b>
	Intercepto	-4.078	- 2.956	-3.630
	Pendiente calibración		0.227	
Hocking 2005	Log-likelihood	-449.350	-428.928	-416.527
	Compañero con infección	2.370	0.877	<b>-0.088</b>
	Numero de compañeros en el último año	0.916	0.339	0.339
	Condomes < 50% de las veces	0.916	0.339	0.339
	Edad <= 25	0.405	0.150	<b>0.942</b>
	<b>Uso de drogas inyectables (&lt;1 año)</b>	1.131		
	Intercepto		-2.747	-3.013
	Pendiente calibración		0.370	
Stergachis	Log-likelihood	-479.367	-429.462	-417.880
	Soltera	1.609	0.377	0.377

Modelos	VARIABLES PREDICTORES*	Coeficientes iniciales	Coeficientes re-calibración	Coeficientes revisión**
	Ectopia cervical	1.308	0.307	0.307
	<b>Duchas en el último 12 años</b>	0.788		
	Nuliparas	0.742	0.174	<b>-0.657</b>
	Edad >=24	0.693	0.163	<b>0.948</b>
	>=2 compañeros sexuales en el último año	0.642	0.151	<b>0.589</b>
	Afroamericano	1.131		
	Intercepto		-2.618	-2.618
	Pendiente calibración		0.235	
Stein 2008	Log-likelihood	-444.005	-427.271	-422.047
	Numero de compañeros en el último año (1)	1.400	0.551	0.551
	Numero de compañeros en el último año (2-50)	2.010	0.791	0.791
	<b>Percepción de riesgo de ITS (Moderado o alto)</b>	1.300		
	<b>No ser estudiante</b>	0.680		
	Edad (18-21)	0.430	0.169	<b>0.779</b>
	Edad (25-26)	0.740	0.291	0.291
	Soltero	1.030	0.406	0.406
	Nunca embarazada	0.570	0.224	<b>-0.416</b>
	<b>No anticoncepción</b>	0.440		
	Intercepto	-6.640	-3.200	-3.113
Pendiente calibración				
Van Duynhoven 1995	Log-likelihood	-406.951	-405.416	
	Edad < 26	0.548	0.446	
	Historia de enfermedad de transmisión sexual	-0.580	-0.472	
	Friabilidad cervical	0.693	0.563	
	Mucopus	0.536	0.436	
	Células guía	0.815	0.663	

Modelos	VARIABLES predictoras*	Coefficientes iniciales	Coefficientes recalibración	Coefficientes revisión**
	Numero de leucocitos $\geq 10$	1.409	1.146	
	Intercepto		-2.685	
	Pendiente calibración		0.813	
Varhoeven 2003	Log-likelihood	-528.734	-425.183	-411.755
	Edad14-17	-0.634	-0.143	-0.143
	Edad 18-22	1.422	0.320	<b>0.963</b>
	Edad 23-27	1.284	0.289	<b>0.925</b>
	Numero de compañeros en el último año (2)	1.790	0.403	<b>0.904</b>
	Numero de compañeros en el último año (3-5)	2.178	0.490	0.490
	Numero de compañeros en el último año ( $\geq 6$ )	2.270	0.511	0.511
	Ninguna anticoncepción	2.438	0.549	<b>0.118</b>
	Disuria	0.883	0.199	<b>-0.441</b>
	Compañero con alguna sintoma urinario	2.050	0.462	0.462
	<b>Sangrado poscoital (Frecuente)</b>	3.048		
	Intercepto	-1.979	-2.717	-2.916
	Pendiente calibración		0.225	

\*Las variables resaltadas en la segunda columna corresponden a aquellas que no se encontraban en el la muestra de validación.

\*\* Los coeficientes resaltados en la última columna son aquellas que requirieron el proceso de revisión.

**Tabla 16** Habilidad discriminativa y medidas de desempeño durante validación y actualización de modelos seleccionados.

Modelo	Parámetros	Validación*	Re-calibración	Revisión
Al-Tayyib 2008	Log-likelihood	-429.355	-426.181	
	Puntaje Brier	0.573	0.086	
	R2	0.027	0.027	
	ROC	0.617	0.617	
	Intercepto	-3.192	0.000	
	Pendiente	0.622	1.000	
Harrison 1985	Log-likelihood	-532.801	-424.728	-418.138
	Puntaje Brier	0.176	0.085	0.084
	R2	0.031	0.031	0.051
	ROC	0.609	0.609	0.651
	Intercepto	-2.029	0.000	-0.099
	Pendiente	0.227	1.000	0.954
Hocking 2005	Log-likelihood	-449.350	-428.928	-416.527
	Puntaje Brier	0.527	0.086	0.084
	R2	0.018	0.018	0.057
	ROC	0.625	0.625	0.651
	Intercepto	-2.747	0.000	0.176
	Pendiente	0.370	1.000	1.082
Stergachis	Log-likelihood	-479.367	-429.462	-417.880
	Puntaje Brier	0.530	0.086	0.084
	R2	0.017	0.017	0.052
	ROC	0.585	0.585	0.636
	Intercepto	-2.616	0.020	-0.130
	Pendiente	0.236	1.007	0.939
Stein 2008	Log-likelihood	-444.005	-427.271	-422.047
	Puntaje Brier	0.092	0.086	0.086
	R2	0.023	0.023	0.033
	ROC	0.607	0.607	0.629
	Intercepto	-0.586	0.000	-0.232
	Pendiente	0.394	1.000	0.950
Van Duynhoven 1995	Log-likelihood	-406.951	-405.416	
	Puntaje Brier	0.315	0.083	
	R2	0.089	0.089	
	ROC	0.704	0.704	
	Intercepto	-2.685	0.000	
	Pendiente	0.813	1.000	

Modelo	Parámetros	Validación*	Re-calibración	Revisión
Verhoeven 2003	Log-likelihood	-528.734	-425.183	-411.755
	Puntaje Brier	0.284	0.086	0.083
	R2	0.030	0.030	0.071
	ROC	0.623	0.623	0.672
	Intercepto	-2.272	0.000	-0.077
	Pendiente	0.225	1.000	0.963

\* Desempeño de los modelos originales en la muestra externa

Todos los modelos tuvieron que ser recalibrados en el intercepto y la pendiente calibración (tabla 16). En general, los modelos presentaron pendientes de calibración entre 0 y 1, demostrando que las probabilidades de los factores predictores necesitaban ser ajustadas. El modelo que más cercanía tuvo en ajuste a las probabilidades observadas en la muestra de validación fue Van Duynhoven, el cual puede observarse por su pendiente de calibración cercana a 1.

En cuanto al intercepto, todos los modelos fueron diferentes de 0, en la misma dirección pero a diferente grado. Esto indica que las probabilidades estimadas por el modelo son demasiado extremas en comparación a las probabilidades observadas (tabla 16). El único modelo que mostró más cercanía a las probabilidades observadas fue Stein (92) debido a que su intercepto calculado en la calibración fue más cercano a 0.

En la validación de los modelos, ellos tuvieron baja desempeño de acuerdo al puntaje de Brier y el estadístico c. Por ejemplo, en el puntaje de Brier, todos los modelos tuvieron un valor superior a 0.09, por lo tanto, los modelos no tienen utilidad para identificar pacientes infectados (Tabla 16).

En discriminación, los modelos tuvieron áreas bajo la curva ROC más bajo que 0.70. El peor modelo fue Al-Tayyib con 0.617 y el mejor modelo fue Van Duyhoven con una AUC de 0.70. Los otros modelos mejoraron en este aspecto posterior a la actualización, pero continuaron presentando pobre exactitud. En el mismo sentido, el  $R^2$  de Nagelkerke incremento después de la actualización dado que esta medida es relativa con el área bajo de la curva ROC (7). Adicional a esto, en la distribución de los desenlaces en cada cuantil que se muestra en la base de la gráfica de calibración, se puede apreciar que entre mejor esta amplitud de ésta distribución mejor es la discriminación (71,77). En general, el intervalo de distribución de los desenlaces fue angosto, por eso, los modelos no tiene buena discriminación entre los pacientes con el desenlace y sin desenlace.

En la re-calibración, los modelos fueron ajustados de acuerdo a la probabilidad observada en la muestra de validación. El efecto del ajuste puede verse en el cambio de los puntajes de Brier en donde los valores fueron inferiores a 0.09. El modelo con más bajo puntaje fue el de Van Duynhoven, con una ligera diferencia de los demás modelos. En este mismo

sentido y como era de esperarse, el logaritmo de la verosimilitud también presentó disminución posterior a la re-calibración.

Finalmente, en la etapa de revisión los modelos mejoraron la capacidad discriminativa y el  $R^2$  de Nagelkerke. Los modelos con mejor desempeño y habilidad discriminativa fueron Van Duynhoven y Verhoeven, pero estas características aún mantienen siendo deficientes. Los modelos de Al-Tayyib y Van Duynhoven no requirieron revisión.

Se puede apreciar en las gráficas de calibración claramente los cambios de la relación de las probabilidades observadas con las probabilidades estimadas antes y después de la revisión (Figura 3). Al ejecutar la calibración y revisión, las probabilidades estimadas tratan de ajustarse a la línea de referencia. A pesar de esto, se observa que los modelos en general no se ajustan adecuadamente a las probabilidades observadas, lo cual explica los resultados obtenidos del puntaje de Brier y  $R^2$  de Nagelkerke.

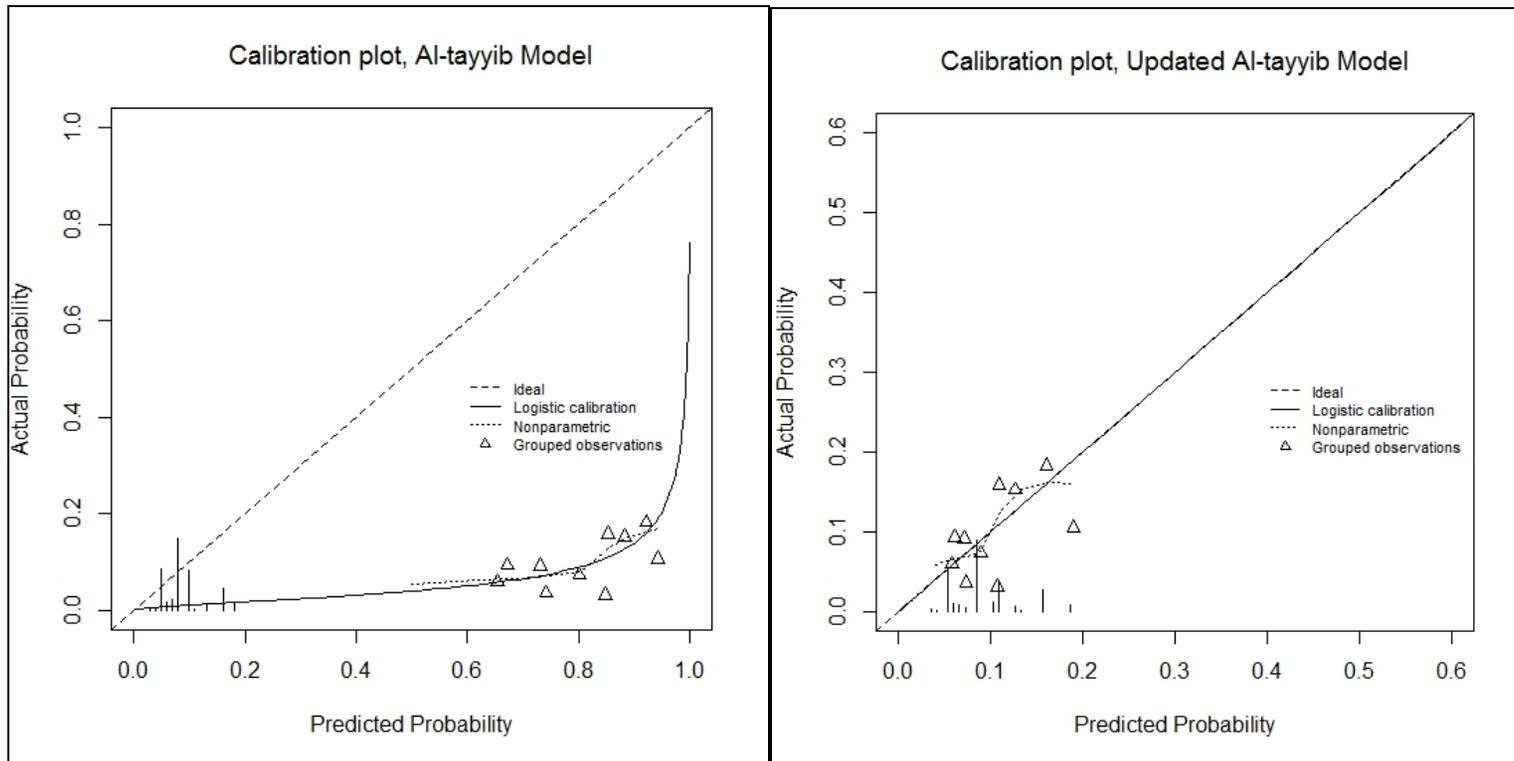
Un hallazgo común fue que la gran mayoría de los modelos tuvieron probabilidad predictiva por encima de la probabilidad observada antes de la calibración, lo cual se explica porque los modelos necesitaron ajustarse en el intercepto. Como se puede notar en la gráfica 3A y 3F, las probabilidades estimadas fueron muy cercanas a las probabilidades observadas en cuanto a su correlación explicando porque estos modelos tuvieron pendientes de calibración más cercanas a 1 que los demás (tabla 16).

Adicionalmente, la distribución de las frecuencia de los eventos fueron localizados entre un rango de 0 – 0.25, es decir, que ningún cuantil presentó un probabilidad observada superior al 25% a pesar de tener más de un factor predictor (parte inferior de las gráficas).

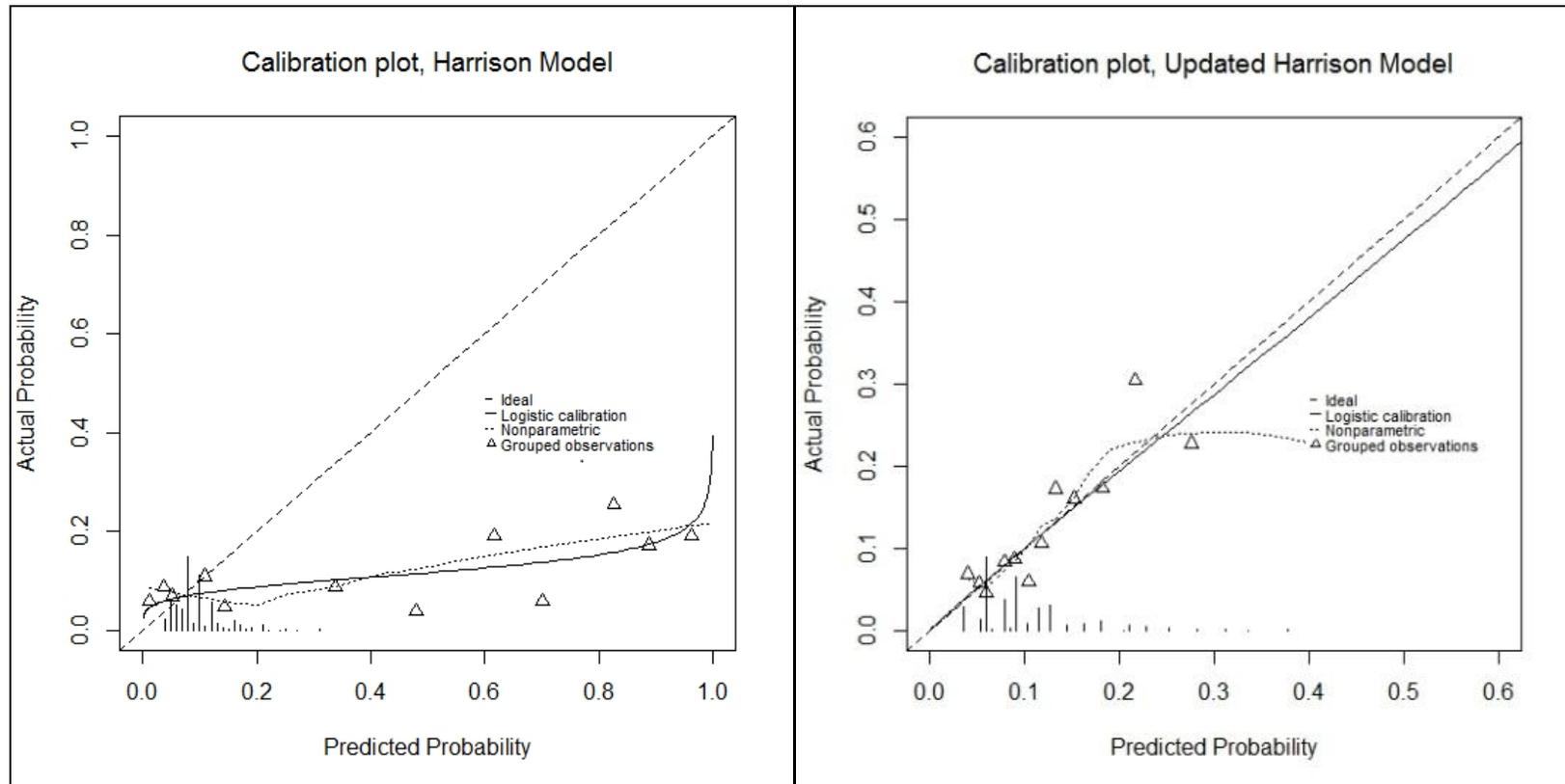
Por otra parte, solamente el modelo de Van Duynhoeven tuvo mejor refinamiento que los otro modelos debido a que demostró una más amplia distribución de los eventos (parte inferior de la gráfica F). Después de la calibración, Verhoeven y Duynhoeven tuvieron mejor distribución de los eventos el cual tiene relación con su resultado en el área bajo la curva ROC. La probabilidad predictiva fue ajustada de acuerdo a la probabilidad observada que fue entre 0-0.3.

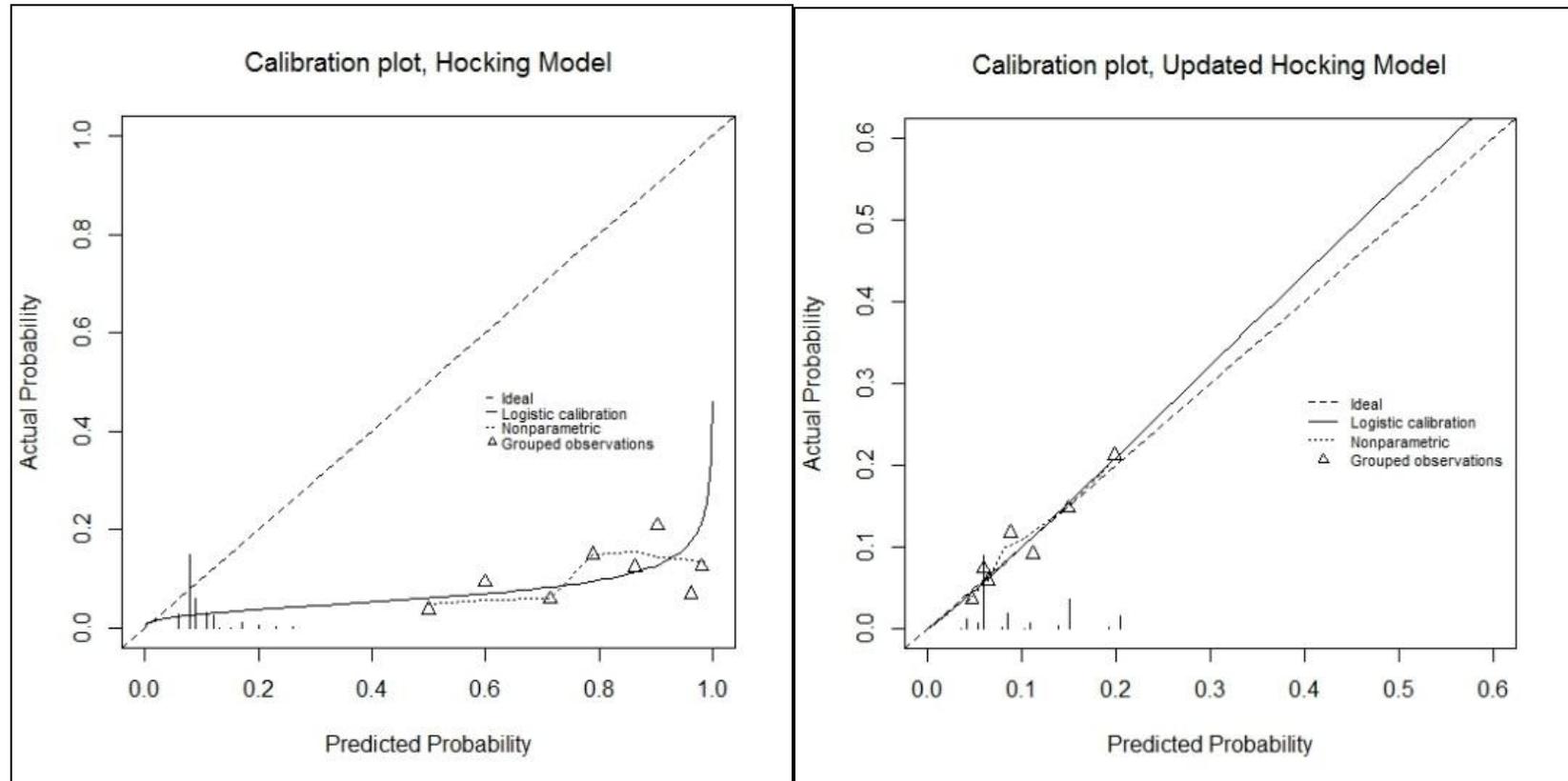
**Figura 3** Comparación entre las gráficas de calibración de los modelos antes y después de la calibración.

## A. Modelo Al-Tayyib

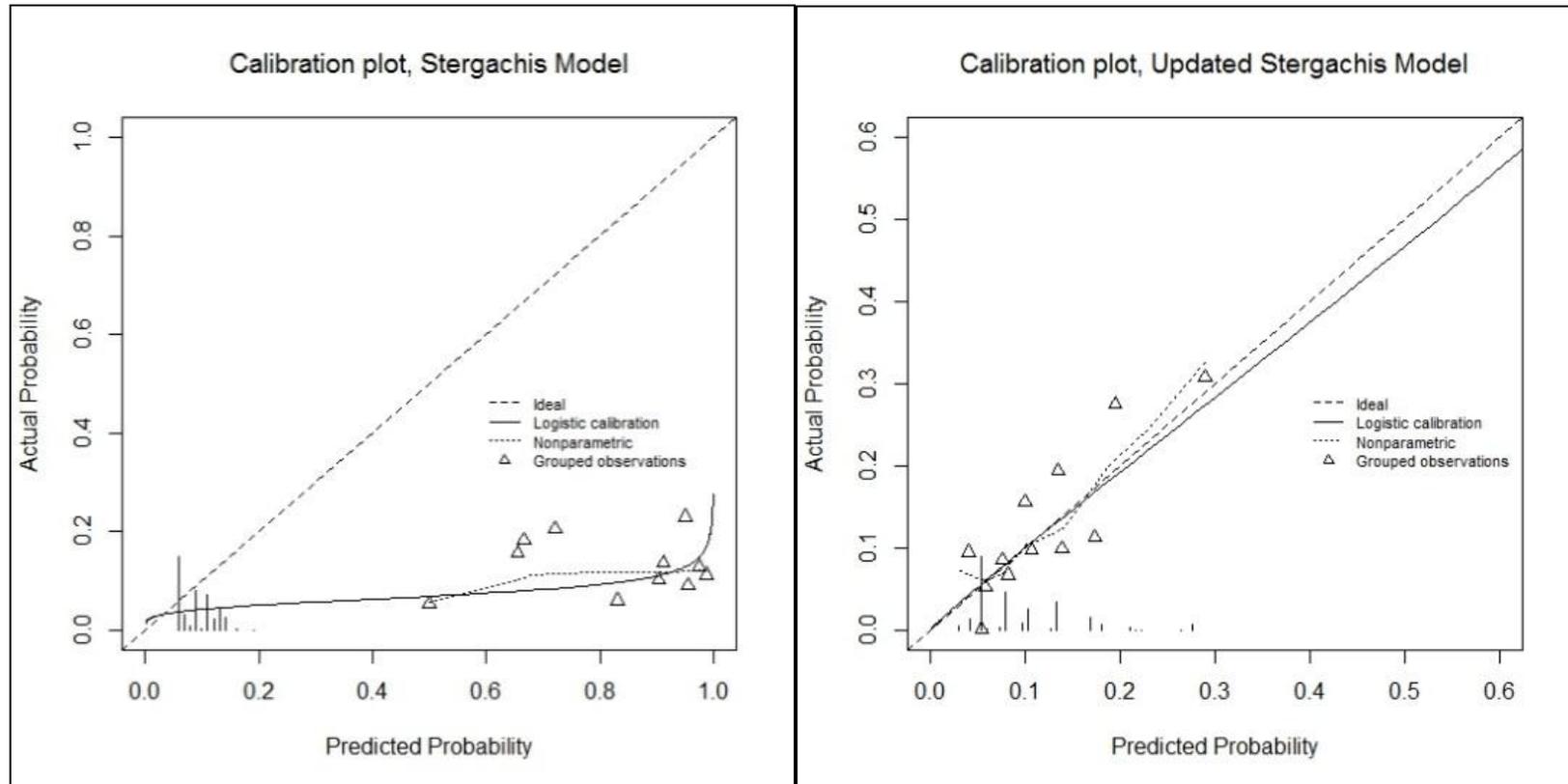


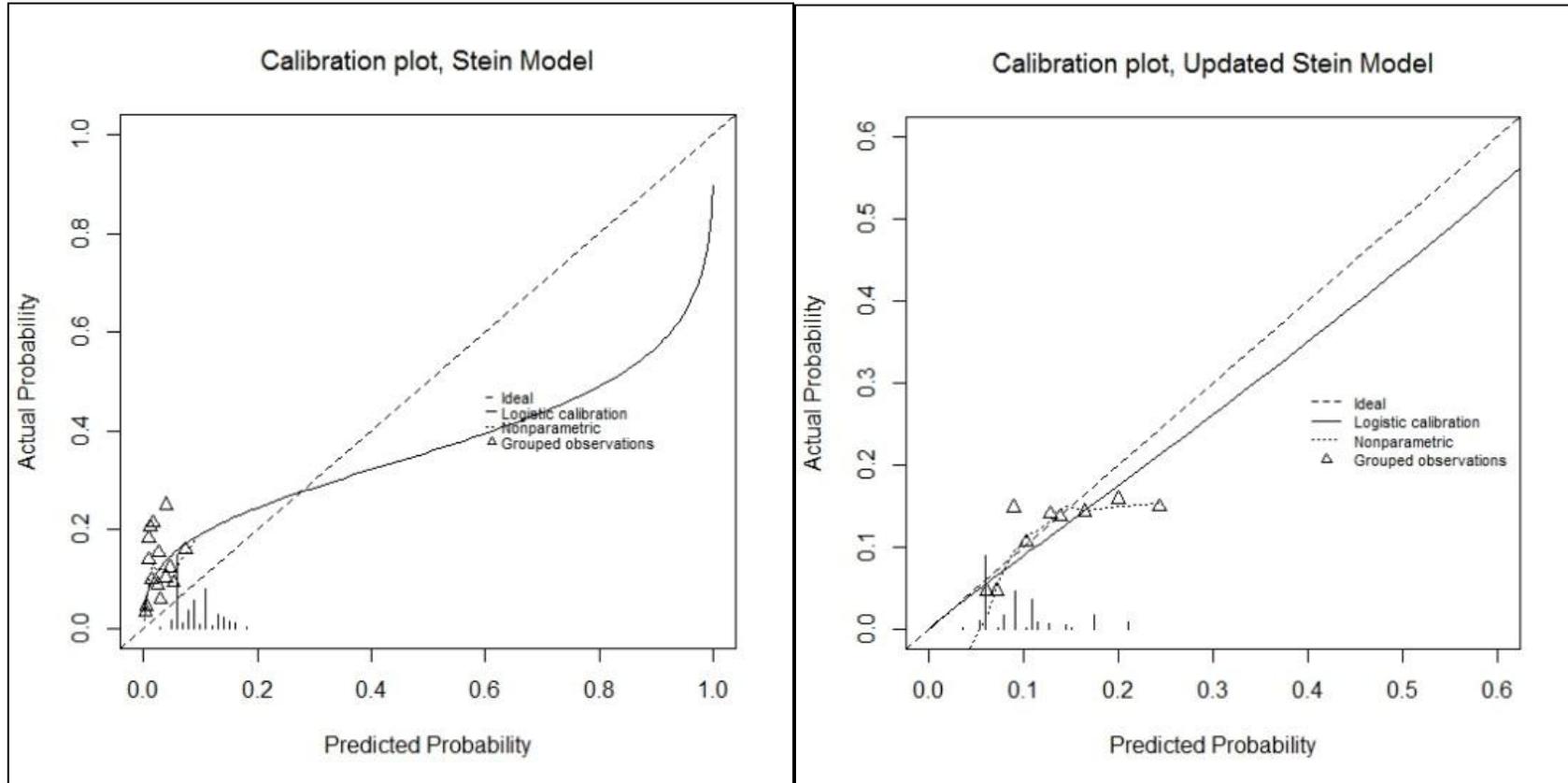
**B. Modelo Harrison**



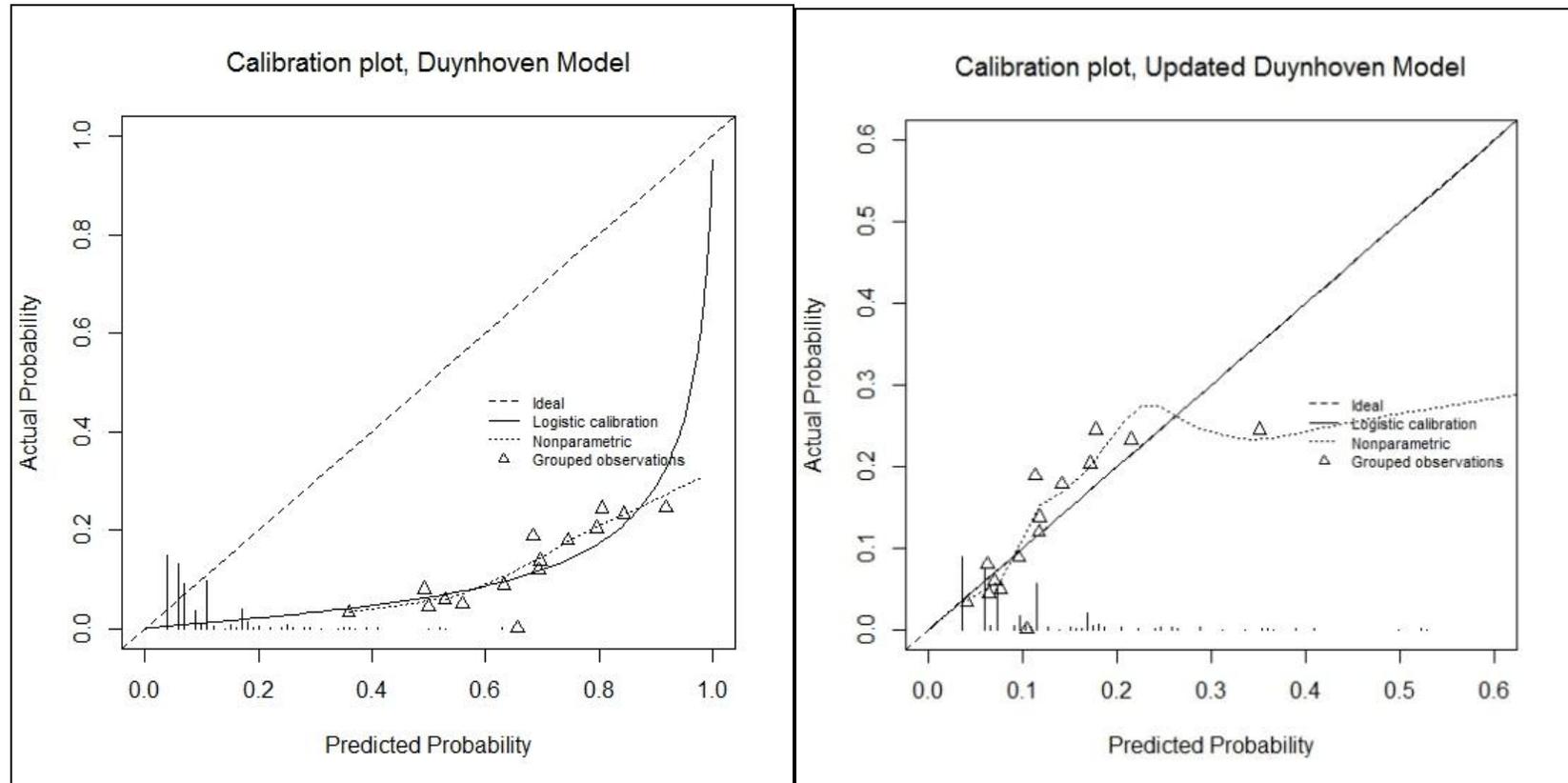
**C. Modelo Hocking**

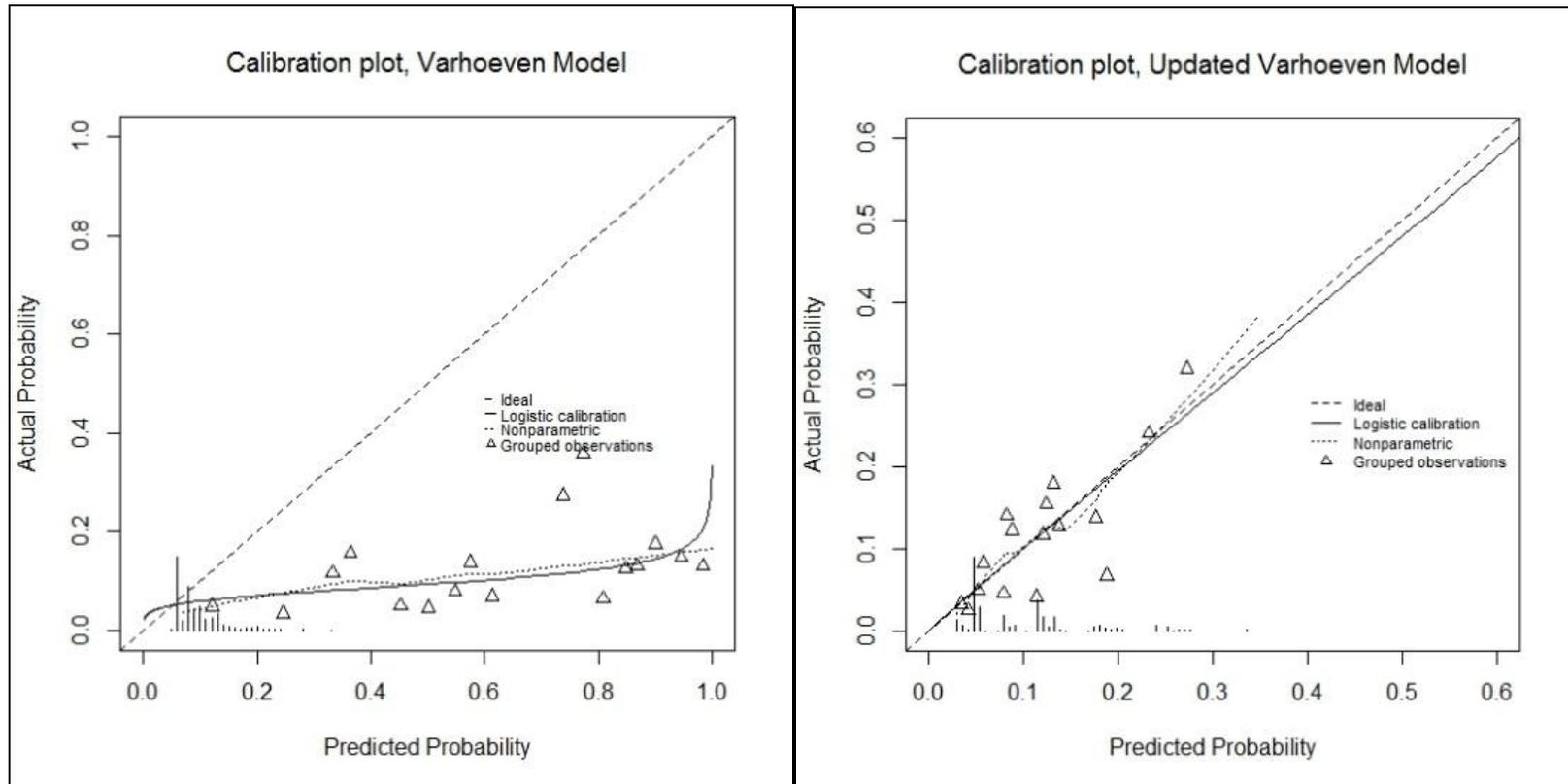
## D. Modelo Stergachis



**E. Modelo Stein**

## F. Modelo Van Duynhoven



**G. Modelo Varhoeven**

### 6.2.3 Método de modelo promediado

Los pesos obtenidos para los modelos usando los datos de calibración y revisión se reportan en la tabla 17. El parámetro BIC más bajo lo presentó el modelo Duynhoven seguido por Al-Tayyib y Hocking. La diferencia de BIC entre el modelo Duynhoven con los demás es bastante amplia que repercute en los pesos finales, esto es debido a que la estimación BIC ésta en base al log verosimilitud y el número de etapas de calibración y revisión donde éste tuvo las cifras más bajas.

El modelo que requirió más etapas en la calibración y revisión fue Varhoeven por lo cual obtuvo un valor de  $u_m$  fue de 10 ( $u_m$  corresponde al factor que se establece según el número de etapas en el proceso de validación y actualización). Entre más variables necesitaba la restimación en el proceso de revisión, el valor  $u_m$  es más alto. Dado que Al-Tayyib y Duynhoven solo requirieron calibración tuvieron valores más bajos de éste parámetro.

**Tabla 17** Resultados de pesos para cada modelo por el método de modelo promediado.

	Al-Tayyib	Harrison	Hocking	Stein	Stergachis	Duynhoven	Varhoven
$l_m$	-426.181	-418.138	-416.527	-422.04675	-417.880	-403.327	-411.755
$u_m (k+1)$	2	5	5	7	6	2	10
N	1376	1376	1376	1376	1376	1376	1376
BIC	866.817	872.411	869.189	894.682	879.122	821.108	895.780
$w_m$	1.2E-10	7.2E-12	3.6E-11	1.1E-16	2.5E-13	1.0E+00	6.1E-17

De acuerdo a estos resultados, los pesos son 0 (Al-tayyib), 0 (Harrison), 0 (Hocking), 0 (Stein), 0 (Stergachis), 1 (Duynhoven) y 0 (Verhoeven). Por lo tanto, el método promediado sugiere que el único modelo en el meta-modelo estaría compuesto por el modelo de Van Duynhoven (Tabla 18).

El modelo Van Duynhoven fue seleccionado por el método de modelo promediado por presentar mejor ajuste y desempeño en la validación externa como representa el log de verosimilitud y el  $u_m$ , en comparación con los otros modelos.

**Tabla 18** Variables y coeficientes para el meta-modelo por el método ponderado.

Variable	Coeficiente
Edad < 26	0.446
Historia de STD	-0.472
Friabilidad cervical	0.563
Mucopus	0.436
Celulas guía	0.663
Number de leucocitos	1.146
Intercepto	-2.685

## 6.2.4 Método de regresión “stacked”

### 6.2.4.1 Generación del meta-modelo

La regresión stacked se obtuvieron los parámetros para cada modelo de predicción, los cuales fueron: 0.1489 (Al-Tayyib), 0.1369 (Harrison), 0.0276 (Hocking), 0 (Stergachis), 0.033 (Stein), 0.1566 (Verhoeven) y 0.6017 (Duynhoven). Según estos parámetros, el modelo de Stergachis es eliminado del meta-modelo. Finalmente, el meta-modelo estaría compuesto por 28 variables donde 17 serían competitivas en el sentido que se repiten en los diferentes modelos. Por lo tanto, estas 17 variables pueden fusionarse como es el caso de la edad, número de compañeros sexuales en los últimos 12 meses, friabilidad, uso de anticonceptivos entre otros.

Como se mencionó anteriormente la variable edad tuvo diferentes definiciones en cada modelo. En algunos fue dicotomizado y en otros fue categorizado. En el meta-modelo esta variable se reajustó como una variable continua con una transformación. Otras variables como el número de células polimorfas y número de compañeros sexuales fueron seleccionadas con coeficientes más altos. Otras variables que medían el mismo atributo, fueron removidos del meta-modelo para evitar colinealidad.

La máxima verosimilitud penalizada fue aplicada en el meta-modelo para reducir el sobreajustamiento. Se obtuvo un factor de penalización de 6.62.

**Tabla 19** Meta-modelo sugerido por la regresión stacked.

Variable	Coficiente	SE
Educación secundaria o menor	0.044	0.168
Disuria	0.060	0.146
Nuevo compañero sexual en el último año	0.041	0.158
Contacto con infección	0.093	0.164
Ninguna anticoncepción	0.111	0.195
Condón < 50% de las veces	0.070	0.129
Historia de STD	-0.307	0.086
Friabilidad cervical	0.367	0.103
Mucopus	0.284	0.08
Células guía	0.431	0.121
Número de compañeros en el año (1)	0.465	0.172
Número de compañeros en el	0.973	0.360

Variable	Coefficiente	SE
año (>2)		
Número de leucocitos < 5	1.610	0.191
Número de leucocitos > 5	3.220	0.382
Edad (Edad <sup>2</sup> )	-0.001	0.0002
Intercepto	-4.768	- 0.158

En la figura 4 se puede observar como sería el modelo expresado en forma de ecuación. Cada variable sería multiplicada con su respectivo coeficiente reportado en la tabla 19.

**Figura 4** Meta-modelo sugerido por el método de la regresión stacked expresado en ecuación

$$\begin{aligned} \text{Log}\left(\frac{P(y)}{1-P(y)}\right) = & \\ & = -4.768 + (0.044 * \text{Educación}) + (0.06 * \text{Disuria}) + (0.041 * \text{Nuevo}) \\ & + (0.093 * \text{Contacto}) + (0.111 * \text{Anticoncepción}) + (0.070 * \text{Condón}) \\ & + (-0.307 * \text{Historia}) + (0.367 * \text{Friabilidad}) + (0.284 * \text{Mucopus}) \\ & + (0.431 * \text{Guía}) + (0.465 * \text{Compañero}(1)) + (0.973 * \text{Compañero} (> 2)) \\ & + (1.610 * \text{Leucocitos}(< 5)) + (3.22 * \text{Leucocitos}(> 5)) \\ & + (-0.001 * \text{Edad}^2) \end{aligned}$$

### 6.2.4.2 Validación interna del meta-modelo

La validación interna mostró satisfactorio desempeño donde el optimismo calculado fue bajo. El puntaje de Brier fue de 0.077, la pendiente y el intercepto fue cerca a línea de referencia de 45° y R<sup>2</sup> fueron relativamente apropiados (Tabla 20) (Figura 5).

Si se compara el puntaje de Brier y R<sup>2</sup> Nagelkerke del meta-modelo con los modelos de predicción anteriores, se observa que estos son inferiores demostrando mejor ajuste a las probabilidades observadas.

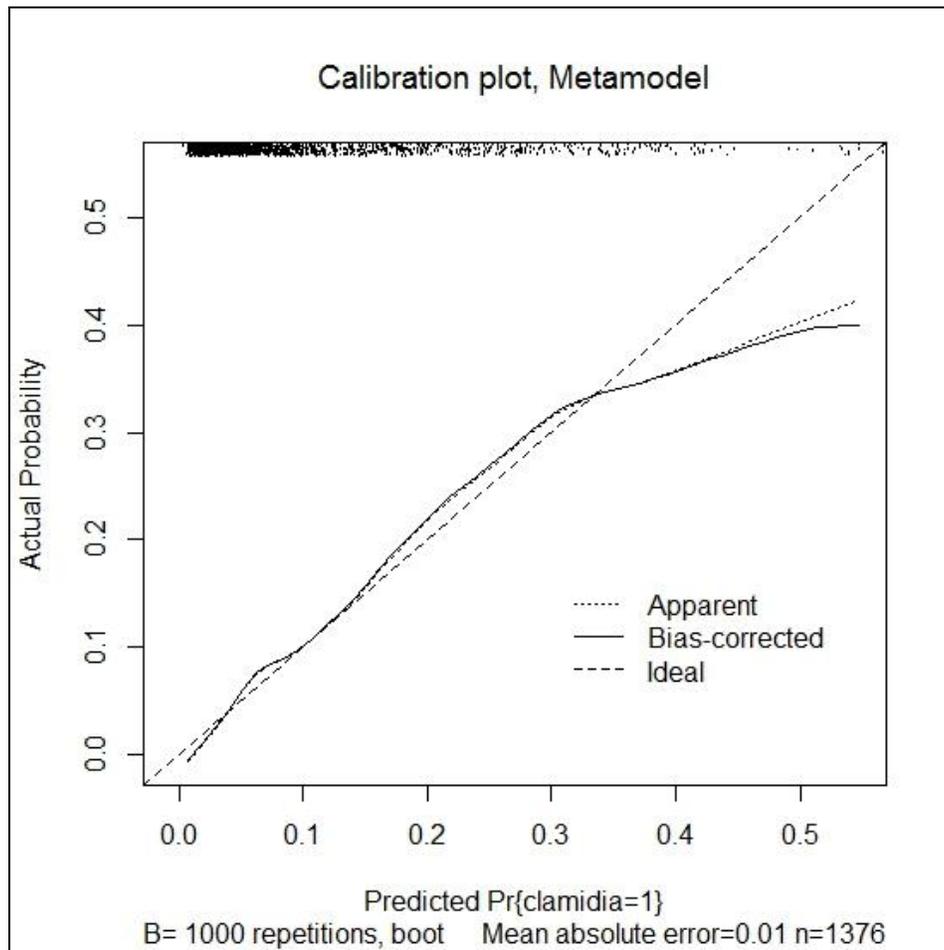
**Tabla 20** Validación interna del Meta-modelo para el método de bootstrapping con 1000 replicaciones.

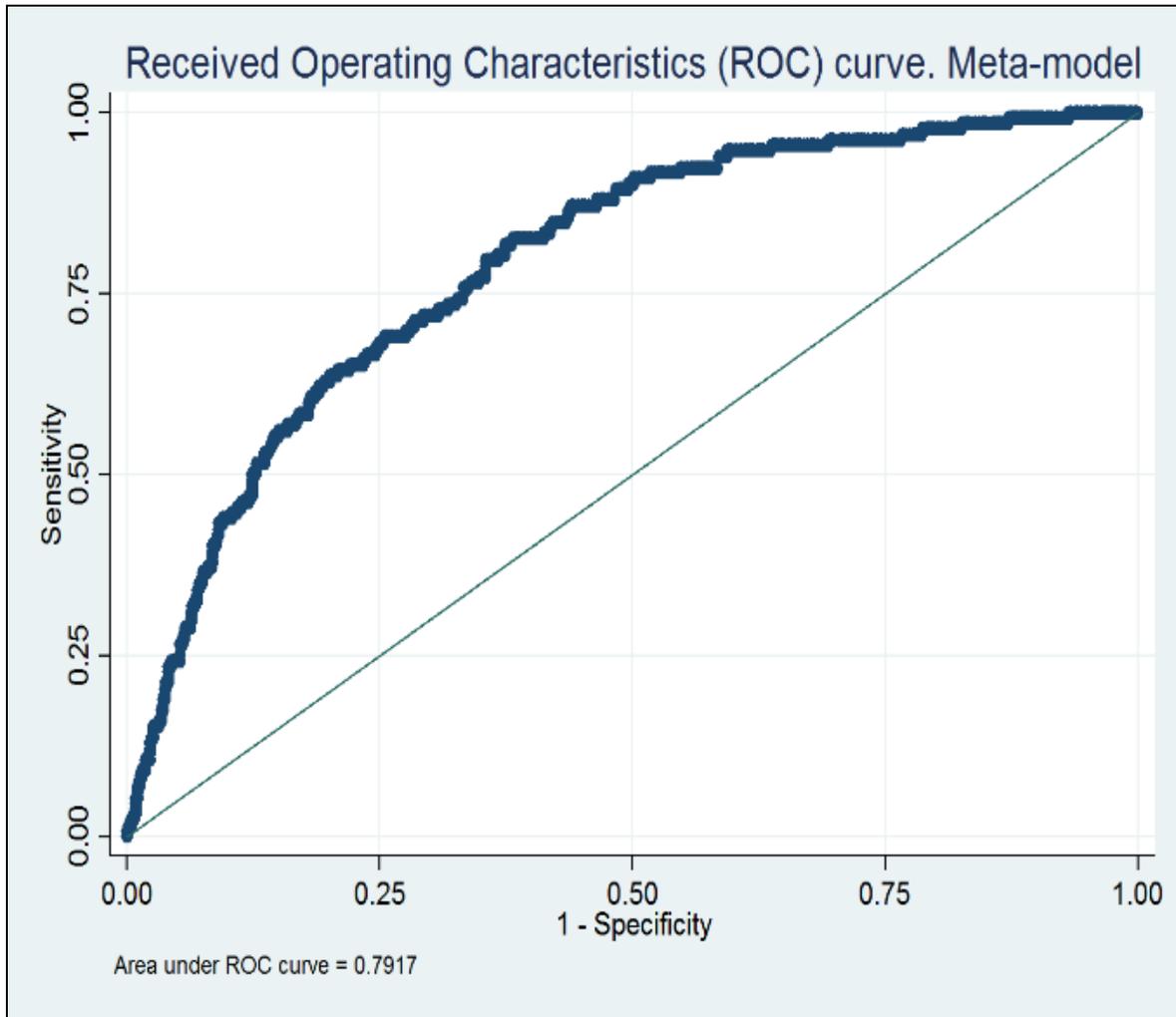
	Indice	Prueba optimismo	Indice corregido
Puntaje Brier	0.077	-0.0003	0.077
R2	0.1972	0.0014	0.1958
Intercepto	0.000	-0.0079	0.008
Pendiente	1.000	-0.0029	1.003

La curva de calibración de la Figura 5 demuestra como el meta-modelo desarrollado por la regresión stacked se ajusta mejor a la línea de referencia que los modelos previos. La estimación del error de máxima calibración fue de 0.0022.

El intervalo de distribución de la frecuencia de los eventos a lo largo de los cuantiles es más amplia que los modelos seleccionados (Parte superior de la figura 5). Esta distribución tendrá efecto en la habilidad discriminativa del meta-modelo y por ende un valor más alto del área bajo la curva ROC (Figura 6). De acuerdo a esta gráfica el área bajo la curva del meta-modelo fue de 0.7917 que es considerado como buena exactitud diagnóstica.

**Figura 5** Curva de calibración de bootstrap para el meta-modelo sugerido por el método de regresión stacked. 1000 repeticiones fueron usadas



**Figura 6** Curva ROC para el meta-modelo del método de regresión de Stacked.

A diferentes puntos de corte, el meta-modelo varía en sensibilidad y especificidad (tabla 21). El punto de corte donde se obtiene el mismo valor de sensibilidad y especificidad fue con una probabilidad de 0.09. Cuando se obtiene una sensibilidad aproximadamente del 90% la especificidad estaría cercana al 51%. Y para una especificidad del 90% la sensibilidad se disminuiría a 43.9%.

**Tabla 21** Características operativas del meta-modelo, el manejo sintomático y prueba rápida para la predicción de infección cervical por *Chlamydia trachomatis* con diferentes puntos de corte.

Prueba diagnóstica	Punto de corte	Sensibilidad (IC 95%)	Especificidad (IC 95%)	VPP (IC 95%)	VPN (IC 95%)	LR+ (IC 95%)	LR- (IC 95%)	ROC
Meta-model	0.05	89.4 (82.8 - 94.1)	51.3 (48.5-54.1)	16.3 (13.7-19.2)	97.9 (96.4-98.8)	1.84 (1.69-1.99)	0.21 (0.13-0.34)	0.792
	0.07	79.5 (71.7-86.1)	63.5 (60.8-66.2)	18.8 (15.6-22.3)	96.7 (95.2-97.8)	2.18 (1.95-2.44)	0.32 (0.23-0.45)	
	0.09	71.2 (62.7-78.8)	71.4 (68.8-73.9)	20.9 (17.2-24.9)	95.9 (94.4-97.1)	2.49 (2.16-2.86)	0.40 (0.31-0.53)	
	0.10	68.2 (59.5-76.0)	74.8 (72.3-77.2)	22.3 (18.4-26.7)	95.7 (94.2-96.9)	2.71 (2.33-3.15)	0.43 (0.33-0.55)	
	0.12	64.4 (55.6-72.5)	78.5 (76.1-80.8)	24.1 (19.8-29.0)	95.4 (93.9-96.6)	3.00 (2.54-3.54)	0.45 (0.36-0.57)	
	0.15	57.6 (48.7-66.1)	83.1 (80.9-85.2)	26.6 (21.5-32.1)	94.9 (93.4-96.1)	3.41 (2.82-4.13)	0.51 (0.42-0.62)	
	0.20	43.9 (35.3- 52.8)	89.8 (88.0 - 91.4)	31.4 (38.6-24.7)	93.8 (92.3-95.1)	4.30 (3.34-5.50)	0.62 (0.54-0.73)	

VPP: Valor predictivo positivo; VPN: Valor predictivo negativo; LR+. Razón de verosimilitud positivo; LR- Razón de verosimilitud negativo

## 6.2.5 Análisis de sensibilidad

Se realizó un análisis de sensibilidad comparando el efecto de las mujeres trabajadoras sexuales en la determinación del meta-modelo bajo el método de la regresión stacked. Para esto se realizó la estimación de la selección de variables y respectivos coeficientes excluyendo a este grupo de mujeres (tabla 22).

**Tabla 22** Meta-modelo desarrollado por la regresión stacked con la muestra de validación sin mujeres trabajadoras sexuales.

Variable	Coeficientes con trabajadoras sexuales	Coeficientes sin trabajadoras sexuales
Educación secundaria o menor	0.044	0.162
Disuria	0.060	0.114
Nuevo compañero sexual en el último año	0.041	0.153
Contacto con infección	0.093	0.248
Ninguna anticoncepción	0.111	0.114
Condón < 50% de las veces	0.070	0.056
Historia de STD	-0.307	-0.290
Friabilidad cervical	0.367	0.347
Mucopus	0.284	0.269
Células guía	0.431	0.408
Numero de compañeros en el año (1)	0.465	0.496
Numero de compañeros en el año (>2)	0.973	1.036
Numero de leucocitos < 5	1.610	1.648
Numero de leucocitos > 5	3.220	3.296
Edad (Edad <sup>2</sup> )	-0.001	-0.001
Intercepto	-4.768	-4.984

Los resultados muestran que las variables que cambiaron sustancialmente en sus coeficientes fueron: haber tenido educación secundaria o menor, disuria, nuevo compañero sexual en el último año, contacto con infección y uso de condón. En todas estas variables hubo aumento en el valor de los coeficientes a excepción de la variable frecuencia de uso de condón.

Las demás variables modificaron levemente en sus coeficientes al determinar el meta-modelo después de usar la base de datos excluyendo a las mujeres trabajadoras sexuales.

## 7. Discusión

### 7.1 Resumen de resultados

Se desarrolló un meta-modelo para el diagnóstico de infección cervical por *Chlamydia trachomatis* a partir de 7 modelos de predicción derivados de la literatura. Este tuvo mejor calibración y habilidad discriminativa que otros modelos de predicción. El modelo fue ajustado con la verosimilitud máxima penalizada y validación interna con bootstrapping para prevenir sobreajustamiento.

La población de los estudios seleccionados fue heterogénea a la muestra de validación. Hubo modelos desarrollados y validados con mujeres jóvenes, jóvenes adultas, sintomáticas, asintomáticas y sintomáticas, mujeres de diferentes continentes y mujeres quienes consultaron en diferentes niveles del cuidado. Estas condiciones permitieron al meta-modelo ser más generalizado para el diagnóstico de la infección cervical debido a que se consideraron las variables predictivas que fueron más significativas en las diferentes poblaciones (62,73).

La revisión sistemática mostró que hay varios modelos de predicción para mujeres no embarazadas con infección cervical en los cuales se presentaron como algoritmos y modelos logísticos. Sin embargo, hay alto riesgo que éstos puedan presentar sobreajustamiento. Primero, muchos de los modelos no fueron validados tanto internamente como externamente (73,74,112). Segundo, el tamaño de la muestra no fue suficiente de acuerdo a la regla de 10 eventos por un candidato predictores (64,74). Tercero, los estudios no reportaron métodos de prevención de sobreajustamiento como el factor shrinkage o la verosimilitud de máxima penalización (83,113). Finalmente, la mayoría de los modelos seleccionaron variables por la estrategia stepwise (114). Por lo tanto, estas reglas de predicción necesitan ser evaluadas cuidadosamente si se espera usar en una nueva población.

Las características de los modelos seleccionados indican posible riesgo de sesgo en la selección de las variables y la estimación de los coeficientes. La mayoría de los estudios usan análisis univariable como primera etapa en la selección de variables lo que podría introducir variables de confusión y rechazar variables que podrían ser significativas en el análisis multivariable a pesar de no ser significativo en el análisis univariado (73,114). Adicionalmente, las variables continuas fueron manejadas como categóricas o dicótomas que pueden producir pérdida de la información y poder de las variables (115). De igual forma, no fue clara la evaluación de interacción en los modelos. Finalmente, el tamaño de la muestra afecta tanto la dirección como la

varianza de los coeficientes cuando estos no son suficientes (64). Todos los factores mencionados anteriormente afectan en el desempeño de los modelos desarrollados y pueden tender a ser inestable.

Respecto a la validación de los modelos que refirieron realizarla, muchos de ellos reportaron la habilidad discriminativa con el área bajo la curva ROC mientras que las medidas de calibración fueron infrecuentemente reportadas. Aunque las medidas de discriminación son importantes, la calibración esta al mismo nivel en un proceso de validación debido a que permite conocer el ajuste del modelo a la base de datos donde fue desarrollado o validado (7,62,77).

Es importante considerar que muchos estudios fueron desarrollados en diferentes periodos antes de esta generación del conocimiento del desarrollo y la validación de modelos de predicción que explican los problemas de reporte en los estudios. De los 25 artículos que se incluyeron a la revisión sistemática, 11 fueron reportados después del año 2000, los cuales corresponden al 44% artículos que tuvieron acceso a este nuevo conocimiento.

En la mayoría de los estudios incluidos, hubo modelos y algoritmos que no tuvieron habilidades discriminativas adecuadas. Solamente, 17 entre modelos y algoritmos reportaron sensibilidad por encima del 80%, pero con baja especificidad. A pesar del hecho que muchos estudios tenían como objetivo desarrollar una prueba para tamizaje y no como diagnóstico, los modelos tuvieron un área bajo la curva ROC cercana al 70% lo cual es considerada como pobre exactitud (116). Estos resultados son debido a que las limitaciones mencionadas anteriormente sobre el desarrollo de los modelos repercuten en el desempeño de estos.

Por otro lado, muchos de los resultados referentes a habilidades discriminativas son sujetos a sobreajustamiento ya que muchos modelos no han sido validados (73). Esto puede notarse en los resultados obtenidos en la validación y actualización de los modelos seleccionados en la muestra externa.

Considerando los resultados de la validación externa de los modelos que reunieron los criterios de inclusión para la agregación, estos presentan baja generalizabilidad en términos de reproducibilidad y transportabilidad. La calibración de los modelos mostró la necesidad de ser ajustados en el intercepto debido a la tendencia sistemática de la probabilidad estimada que era más alta a la probabilidad observada. A pesar de la validación y actualización de los modelos, la gráfica de calibración de los modelos no se ajustó debidamente a la línea de referencia. Los modelos donde se puede apreciar claramente este comportamiento son de Al-Tayyib, Stergachis, Stein y Varhoeven.

En la etapa de revisión, solamente los modelos de Al-Tayyib y Duynhoven no requirieron de esta debido a que las variables predictores y sus coeficientes se

ajustan al comportamiento de la muestra de validación. Los demás modelos requirieron ajuste de algunas variables como por ejemplo la edad, número de compañeros sexuales, nulíparas y anticoncepción. Esto principalmente puede ser causado por el efecto caso-mixto donde hay diferencias entre las muestras de desarrollo del modelo y de validación (117).

Se usaron diferentes métodos para la agregación de los modelos. El método de modelo promediado seleccionó solamente el modelo de Van Duyhoeven por tener menor error durante la validación externa y actualización. Los otros modelos no fueron seleccionados por que tuvieron bajo desempeño y necesitaban ser actualizados ampliamente. En este método, la penalización depende de los resultados del logaritmo de la verosimilitud y si ellos requieren más pasos en la revisión (34). Por ejemplo, a pesar que el modelo de Al-Tayyib no requirió revisión sino solamente calibración, este tuvo alta valor en el logaritmo de verosimilitud mientras que el modelo de Van Duynhoven no requirió revisión y tuvo un bajo valor de logaritmo de verosimilitud. Este método de agregación es criticado en el sentido que es un método bastante sensible en las variaciones y por ende trata de seleccionar el mejor modelo más que la agregación (34). Por su fórmula de cálculo, en donde las diferencias se realizan en forma cuadrática, la diferencia de aporte entre los modelos es más distante entre uno y el otro.

Por otro lado, el método de regresión stacked solamente descartó el modelo de Stergachis mientras que los otros modelos se mantuvieron en el meta-modelo. En este caso se obtuvo un meta-modelo de 15 variables. Estas corresponden a las factores predictores más frecuentes de los 31 reglas de predicción incluidas en la revisión sistemática (tabla 10). El coeficiente de cada variable depende de si se encuentra contenido en otros modelos, los coeficientes iniciales y el parámetro estimado con la regresión stacked. Es importante resaltar que este método busca la mejor combinación de los diferentes modelos de acuerdo a la población donde se va aplicar.

El desempeño de la validación interna del meta-modelo por la regresión stacked mostró bajo sobreajustamiento después de 1000 replicaciones ya que las medidas de calibración y discriminación fueron corregidas ligeramente. De acuerdo con estos resultados el meta-modelo tuvo mejores resultados de los otros modelos. El estadístico  $c$  o área bajo la curva de 0.79, el cual es considerado como buena exactitud. Así mismo, el puntaje de Brier y el  $R^2$  de Nagelkerke fue inferior a los demás modelos. También, la calibración mostró ajustarse mejor a la línea de referencia con bajo error absoluto de calibración. Cabe notar, que este modelo fue agregado basado a la base de datos de las mujeres de Colombia lo que ayudó a obtener estos resultados. No obstante, se intentó evitar sobreajustamiento con el cálculo de coeficientes penalizado.

La regresión stacked trabaja usando algoritmos de aprendizaje donde los modelos son convertidos en predictores para derivar el meta-modelo (34,118). El proceso es basado en la validación cruzada. De acuerdo a Clark (119) y Debray (34), la regresión

stacked es un método más robusto dado que el método de modelo ponderado es más exigente con el error o variaciones de los modelos. Esto puede notarse en los resultados del método del modelo ponderado donde la diferencia entre modelo Van Duynhoven y los otros modelos fue muy amplia. En cambio, la regresión stacked a pesar de esta diferencia entre los modelos fueron adicionados al meta-modelo.

Si se observan los pesos calculados por Stacked de cada modelo, se puede notar que el modelo de mayor peso sigue siendo Van Duynhoven con 0.60, seguidos por los modelos Harrison, Verhoeven y Al-Tayyib. Los modelos de menor aporte al meta-modelo fueron Stein y Hocking.

Observando los bajos desempeños de los modelos seleccionados y el meta-modelo desarrollado, la cervicitis producida por *Chlamydia trachomatis* muestra ser una infección de difícil predicción si se compara con otros escenarios clínicos, como por ejemplo en predicción de trombosis venosa donde se han obtenido modelos con áreas bajo la curva cercanos al 0.90.

### **Aplicabilidad del modelo**

Es necesario seleccionar un punto de corte del modelo dependiendo del objetivo del mismo. En orden de seleccionar el correcto punto de corte se debería contar con expertos clínicos los cuales evalúan las implicaciones en una toma de decisiones y que es más favorable para la paciente. En un caso ideal, la principal recomendación es realizar un análisis de impacto en donde se evalúa el impacto de la implementación de la regla de predicción en decisiones clínicas (120).

En el marco de las infecciones cervicales por clamidia, hay que tener en cuenta dos situaciones críticas. Primero, las mujeres quienes tienen cervicitis pero no ha sido diagnosticadas (falso negativo), quienes pueden sufrir secuelas como desarrollo de enfermedad pélvica inflamatoria, infertilidad y embarazo ectópico. Segundo, las mujeres quienes no tienen infección cervical pero fueron diagnosticadas (falsos positivos) deben considerarse el riesgo de desarrollo de eventos adversos, resistencia de antibióticos y el costo del tratamiento. Por lo tanto, es necesario de un estudio de análisis de impacto en donde se tendría en cuenta estas dos situaciones el cual dependería de la probabilidad que ocurra eventos y el impacto que pueda producir.

A pesar de no contar con estos estudios de análisis de impacto, se puede comparar la exactitud diagnóstica del meta-modelo con el algoritmo sindrómico de la Organización Mundial de Salud (OMS) (Sensibilidad 13.7 (IC95% 8.57-20.4) y especificidad 91 (89.3-92.5)) las cuales fueron aplicadas a la misma muestra en el estudio realizado por Rodríguez A (5). Para el algoritmo de la OMS, se encuentra que el meta-modelo en todos sus puntos de corte, la razón de verosimilitud positiva es más alta y la razón de verosimilitud negativa es más baja. Esto significa que el meta-modelo muestra un mejor desempeño que el algoritmo de OMS. Sin contar el impacto social que puede tener este modelo.

La probabilidad de clasificar mejor a un paciente enfermo con respecto a un sano es más alta con el meta-modelo en comparación con el algoritmo de la OMS y la prueba rápida, como lo demuestra el área bajo la curva ROC. Adicionalmente, en la práctica clínica el meta-modelo presenta las mismas ventajas de fácil aplicabilidad que el algoritmo de la OMS en donde se puede usar en los diferentes niveles de atención ya que las variables que alimentan el meta-modelo son de fácil medición durante la consulta. Por lo tanto, aseguraría un mayor cubrimiento a las pacientes con ésta infección en todo del territorio colombiano. Sin contar la fácil implementación por los recursos que se requieren para ello.

### ***Calidad de la evidencia***

Se usó la herramienta PROBAST en orden de evaluar la calidad de evidencia de los estudios seleccionados. En general, los estudios no tuvieron buena calidad por alto riesgo de sesgo. El principal problema en la calidad perteneció a los dominios de la población, tamaño de la muestra y el análisis de los datos.

El dominio de la población, la mayoría de los estudios fue calificado como no claro por el predominio de falta de información en los criterios de selección de la población y proceso de reclutamiento. No fue posible juzgar si se presentó sesgo de selección en los estudios incluidos.

En el dominio de tamaño de la muestra y flujo, el principal problema en los estudios consistió en el cumplimiento de la relación de 10 eventos por candidato. En este dominio muchos de los artículos tuvieron alto riesgo de sesgo en el desarrollo de los modelos o su desempeño. Como se mencionó anteriormente, este puede incrementar el riesgo de sobreajustamiento y la estimación de los coeficientes.

Finalmente, en dominio de análisis hubo alto riesgo porque los modelos no fueron validados o correctamente validados. Así, hubo alto riesgo de interpretación de los resultados y de las medidas en los modelos.

### ***Potenciales sesgos***

El tamaño de la muestra de este estudio fue de 133 desenlaces y 1248 mujeres sin desenlace. De acuerdo con Steyerberg (33), y Vergouwe (95), en esta clase de estudios con métodos de actualización de modelos de predicción es recomendable un tamaño de muestra superior a 1000-2000 pacientes (con y sin desenlace) en combinación con el factor shrinkage. Aunque, el tamaño de muestra no fue deficiente, los resultados serían más confiables si se hubiera tenido un tamaño de muestra mas grande. Sin embargo, se ejecutó el cálculo con la máxima verosimilitud penalizada en pro de prevenir el sobreajustamiento que es un método similar al factor shrinkage.

Un importante sesgo de selección fue la no evaluación de todos los modelos. La principal razón de no selección fue la incompleta información de variables predictoras en la muestra de validación y del reporte de coeficientes correspondientes. Tampoco

se puede realizar un análisis del impacto de este sesgo ya que por ser probabilidades condicionales es difícil utilizar un modelo incompleto.

Adicionalmente, pudo haber sesgo de publicación por no utilizar artículos no publicados en la revisión sistemática. Sin embargo, en la revisión sistemática, no se estableció filtros sobre lenguaje o año de publicación.

Finalmente, el meta-análisis tiene incertidumbres sobre su desempeño en mujeres trabajadoras sexuales. Aunque, hubo trabajadoras sexuales en la muestra de validación, su prevalencia fue de 8.7%. Se realizó un análisis de sensibilidad en el desarrollo del meta-modelo con y sin el grupo de mujeres trabajadoras sexuales en donde se encontró ligeros cambios en los coeficientes. No se encontraron variación en las habilidades discriminativas. Sin embargo, a pesar de estos resultados, es probable que esta población presente otro comportamiento en donde requiera otro tipo de variables predictoras y con otra fuerza predictora. No obstante, en esta muestra no pudo notarse diferencia por el bajo porcentaje de trabajadoras sexuales.

Las variables que presentaron variaciones en el análisis de sensibilidad mostraron que las trabajadoras sexuales diluyen la fuerza de los predictores ya que al excluir este grupo, el coeficiente aumenta. De modo contrario, ocurrió con la variable frecuencia de uso de condón en donde la variable se hizo más débil. La principal razón de estos cambios corresponde a que estos factores predictivos son frecuentes en las mujeres sexuales que no presentan la infección de clamidia como por ejemplo número de compañeros sexuales en el último año.

#### **Acuerdos y desacuerdos con otros estudios**

Los hallazgos sobre el desempeño de los modelos incluidos en las revisiones sistemáticas concuerdan con la revisión de la literatura realizado por Sloan (121) en el 2000 en donde tenía como objetivo estudiar las características operativas de las diferentes herramientas de diagnósticas para las infecciones por clamidia y gonorrea. Sloan concluyó que las herramientas desarrolladas hasta la época tenían pobre desempeño. Sin embargo, este estudio no fue desarrollado en forma sistemática y tampoco se tuvieron criterios claros en la selección de los estudios llevando a posibles sesgos de selección.

Es importante resaltar que la revisión sistemática realizada trato de evitar la selección de modelos con sesgos de clasificación al excluir los estudios que diagnosticaron infección clamidial con otras pruebas diagnósticas diferentes al patrón de oro. Adicionalmente, mantuvo en cuenta la perspectiva de evaluación de sesgos en el desarrollo y validación de reglas de predicción clínica para este tipo de escenario clínico.

Finalmente, este es uno de los primeros estudios de agregación de modelos predictivos para infecciones por clamidia debido al reciente desarrollo de los métodos descritos por el doctor Debray.

***Implicaciones en la práctica e investigación***

Se desarrolló una nueva herramienta diagnóstica para cervicitis por *Chlamydia trachomatis* enfocada a la población colombiana que mostro superioridad en las características operativas al algoritmo de la OMS y ligeramente mejor que las pruebas rápidas. Este meta-modelo puede ser un nueva prueba diagnóstica de bajo costo, fácil uso en todos los niveles de atención y fácil implementación en países emergentes y no desarrollados.

Como ya se mencionó anteriormente, se consideraron estudios que usaron pruebas de referencia para diagnóstica las infecciones clamidiales, por ende, se evitó sesgo en la selección de modelos con predictores y correspondientes coeficientes sesgados garantizando mayor estabilidad en el proceso de agregación.

Adicionalmente, se utilizó una nueva herramienta en el campo de modelos de predicción que permitirá utilizar la información de modelos previamente desarrollados para crear meta-modelos que serían más robustos y generalizables gracias a la heterogeneidad de las poblaciones donde fueron creados. Con estos nuevos modelos, sencillamente se necesitaría ajustarlos a la nueva población en donde se quiera utilizar.

## 8. Conclusión

Se desarrolló un modelo de predicción diagnóstica de infección cervical por *Chlamydia trachomatis* con características operativas superiores de otros modelos previamente desarrollados y el diagnóstico sindrómico planteado por la OMS. Este modelo puede tener alto impacto en salud pública ya que permite un tratamiento oportuno de la infección deteniendo la cadena de transmisión y evitando el desarrollo de secuelas en las pacientes infectadas. Por otro lado, la regresión stacked mostró ser un método eficiente para el meta-análisis de modelos de predicción con mejoramiento de las características operativas de los modelos de origen.

## Bibliografía

1. World Health Organization. Sexually transmitted infections (STIs) [Internet]. Fact Sheet N° 110. 2013 [cited 2014 Feb 12]. Available from: <http://www.who.int/mediacentre/factsheets/fs110/en/>
2. Breu F, Guggenbichler S, Wollmann J. Global incidence and prevalence of selected curable sexually transmitted infections - 2008. WHO Organ. 2008;
3. World Health Organization. Global prevalence and incidence of selected curable sexually transmitted infections: overviews and estimates. 2001.
4. Harrell FE, Margolis PA, Gove S, Mason KE, Mulholland EK, Lehmann D, et al. Development of a clinical prediction model for an ordinal outcome: the World Health Organization Multicentre Study of Clinical Signs and Etiological agents of Pneumonia, Sepsis and Meningitis in Young Infants. WHO/ARI Young Infant Multicentre Study Group. *Stat Med*. 1998;17(8):909–44.
5. Rodríguez A. Validez del diagnóstico sintomático de cervicitis y flujo vaginal en mujeres en edad reproductiva, Bogotá-Colombia 2010. Universidad Nacional de Colombia; 2013.
6. Greer L, Wendel GD. Rapid Diagnostic Methods in Sexually Transmitted Infections. *Infectious Disease Clinics of North America*. 2008. p. 601–17.
7. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Book. New York: Springer; 2009.
8. Miller WC. Screening for chlamydial infection. A model program based on prevalence. *Sex Transm Dis*. 1998;25(4):201–10.
9. World Health Organization. Infecciones de transmisión sexual y otras infecciones del tracto reproductivo. Una guía para la práctica básica. Primera. Geneva: World Health Organization; 2005.
10. European Centre for Disease Prevention and Control. Reporting on 2011 surveillance data and 2012 epidemic intelligence data. *Annual Epidemiological Report 2013*. 2013.
11. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2012. Atlanta, Georgia; 2014.
12. Ángel Müller E, Rodríguez A, Núñez Forero LM, Moyano LF. Prevalencia y factores asociados a la infección por *C. trachomatis*, *N. gonorrhoeae*, *t. vaginalis*, *C. albicans*, sífilis, VIH y vaginosis bacteriana en mujeres con síntomas de infección vaginal en tres sitios de atención de Bogotá, Colombia, 2010. *Rev Colomb Obstet Ginecol*. 2012;63(1):14–24.
13. Centers for Disease Control and Prevention. Recommendations for the prevention and management of *Chlamydia trachomatis* infections, 1993. *MMWR*. 1993;42(RR-12):1–39.
14. Gottlieb SL, Xu F, Brunham RC. Screening and treating *Chlamydia trachomatis* genital infection to prevent pelvic inflammatory disease: interpretation of findings from randomized controlled trials. *Sex Transm Dis*. 2013;40(2):97–102.
15. Peipert JF. Clinical practice. Genital chlamydial infections. *N Engl J Med*. 2003 Dec 18;349(25):2424–30.
16. Sweet RL, Landers D V, Walker C, Schachter J. *Chlamydia trachomatis* infection and pregnancy outcome. *Am J Obstet Gynecol*. 1987;156(4):824–33.
17. Hammerschlag MR. Chlamydial and gonococcal infections in infants and children. *Clin Infect Dis*. 2011;53 Suppl 3:S99–102.

18. Schachter J, Grossman M, Sweet RL, Holt J, Jordan C, Bishop E. Prospective study of perinatal transmission of *Chlamydia trachomatis*. *Jama*. 1986;255(24):3374–7.
19. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sex Transm Infect*. 1999;75(1):3–17.
20. Mukenge-Tshibaka L, Alary M, Lowndes CM, Van Dyck E, Guédou A, Geraldo N, et al. Syndromic versus laboratory-based diagnosis of cervical infections among female sex workers in Benin: implications of nonattendance for return visits. *Sex Transm Dis*. 2002;29(6):324–30.
21. Peeling RW, Mabey D, Herring A, Hook EW. Why do we need quality-assured diagnostic tests for sexually transmitted infections? *Nat Rev Microbiol*. 2006;4(12 Suppl):S7–S19.
22. Schwebke JR, Sadler R, Sutton JM, Hook EW. Positive screening tests for gonorrhea and chlamydial infection fail to lead consistently to treatment of patients attending a sexually transmitted disease clinic. *Sex Transm Dis*. 1997;24(4):181–4.
23. Kamali A, Quigley M, Nakiyingi J, Kinsman J, Kengeya-Kayondo J, Gopal R, et al. Syndromic management of sexually-transmitted infections and behaviour change interventions on transmission of HIV-1 in rural Uganda: A community randomised trial. *Lancet*. 2003;361(9358):645–52.
24. Alary M, Baganizi E, Guèdèmè A, Padonou F, Davo N, Adjovi C, et al. Evaluation of clinical algorithms for the diagnosis of gonococcal and chlamydial infections among men with urethral discharge or dysuria and women with vaginal discharge in Benin. *Sex Transm Infect*. 1998;74 Suppl 1:S44–S49.
25. Pettifor A, Walsh J, Wilkins V, Raghunathan P. How effective is syndromic management of STDs?: A review of current studies. *Sex Transm Dis*. 2000 Aug;27(7):371–85.
26. Pépin J, Deslandes S, Khonde N, Kintin DF, Diakité S, Sylla M, et al. Low prevalence of cervical infections in women with vaginal discharge in west Africa: implications for syndromic management. *Sex Transm Infect*. 2004;80(3):230–5.
27. Tyndall MW, Kidula N, Sande J, Ombette J, Temmerman M. Predicting *Neisseria gonorrhoeae* and *Chlamydia trachomatis* infection using risk scores, physical examination, microscopy, and leukocyte esterase urine dipsticks among asymptomatic women attending a family planning clinic in Kenya. *Sex Transm Dis*. 1999;26(8):476–82.
28. Rugpao S, Rungruengthanakit K, Werawatanakul Y, Sinchai W, Ruengkris T, Lamlertkittikul S, et al. Risk factors and algorithms for chlamydial and gonococcal cervical infections in women attending family planning clinics in Thailand. *J Obstet Gynaecol Res*. 2010;36(1):147–53.
29. Harrison HR, Costin M, Meder JB, Bownds LM, Sim DA, Lewis M, et al. Cervical *Chlamydia trachomatis* infection in university women: relationship to history, contraception, ectopy, and cervicitis. *Am J Obstet Gynecol*. 1985;153(3):244–51.
30. Stergachis A, Scholes D, Heidrich FE, Sherer DM, Holmes KK, Stamm WE. Selective screening for *Chlamydia trachomatis* infection in a primary care population of women. *Am J Epidemiol*. 1993;138(3):143–53.
31. Verhoeven V, Avonts D, Meheus A, Goossens H, Ieven M, Chapelle S, et al. Chlamydial infection: an accurate model for opportunistic screening in general practice. *Sex Transm Infect*. 2003;79(4):313–7.
32. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. *PLoS Medicine*. 2012;9(5):e1001221.
33. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, Eijkemans MJC, Habbema JDF. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567–86.
34. Debray T. Meta-analysis of clinical prediction models. Utrecht, Netherland: Utrecht University; 2013.
35. Manavi K. A review on infection with *Chlamydia trachomatis*. *Best Pract Res Clin Obstet Gynaecol*. 2006;20(6):941–51.

36. Sweet RL, Gibbs RS. *Infectious Diseases of the Female Genital Tract*. Fifth Edit. Philadelphia: PA Wolters Kluwer/Lippincott Williams & Williams; 2009.
37. Faro S, Soper DE. *Enfermedades Infecciosas En La Mujer*. Mexico: McGraw Hill; 2002.
38. Da Ros CT, Schmitt C da S. Global epidemiology of sexually transmitted diseases. *Asian J Androl*. 2008;10(1):110–4.
39. Mérida SR, Iván R-PA, Lucía O-OO. Prevalencia de Chlamydia Trachomatis detectada por reacción en cadena de la polimerasa en un grupo de mujeres jóvenes sintomáticas y asintomáticas en Bogotá, Colombia. *Rev Colomb Obs Ginecol*. 2006;57(3):171–81.
40. Franceschi S, Smith JS, van den Brule A, Herrero R, Arslan A, Anh P-T-H, et al. Cervical infection with Chlamydia trachomatis and Neisseria gonorrhoeae in women from ten areas in four continents. A cross-sectional study. *Sex Transm Dis*. 2007;34(8):563–9.
41. Jaschek G, Gaydos CA, Welsh LE, Quinn TC. Direct detection of Chlamydia trachomatis in urine specimens from symptomatic and asymptomatic men by using a rapid polymerase chain reaction assay. *J Clin Microbiol*. 1993;31(5):1209–12.
42. Harindra V, Underhill G, Tobin JM. Screening for genital chlamydia infection: DNA amplification techniques should be the test of choice. *Int J STD AIDS*. 2003;14(11):723–6.
43. Herring A, Ballard R, Mabey D, Peeling RW. Evaluation of rapid diagnostic tests: chlamydia and gonorrhoea. *Nat Rev Microbiol*. 2006;4(12 Suppl):S41–8.
44. Tuuminen T, Palomäki P, Paavonen J. The use of serologic tests for the diagnosis of chlamydial infections. *J Microbiol Methods*. 2000;42(3):265–79.
45. Rabenau HF, Köhler E, Peters M, Doerr HW, Weber B. Low correlation of serology with detection of Chlamydia trachomatis by ligase chain reaction and antigen EIA. *Infection*. 2000;28(2):97–102.
46. Morré SA, Munk C, Persson K, Krüger-Kjaer S, van Dijk R, Meijer CJLM, et al. Comparison of three commercially available peptide-based immunoglobulin G (IgG) and IgA assays to microimmunofluorescence assay for detection of Chlamydia trachomatis antibodies. *J Clin Microbiol*. 2002;40(2):584–7.
47. Verkooyen RP, Peeters MF, van Rijsoort-Vos JH, van der Meijden WI, Mouton JW. Sensitivity and specificity of three new commercially available Chlamydia trachomatis tests. *Int J STD AIDS*. 2002;13 Suppl 2:23–5.
48. Bauwens JE, Clark AM, Stamm WE. Diagnosis of Chlamydia trachomatis endocervical infections by a commercial polymerase chain reaction assay. *J Clin Microbiol*. 1993;31(11):3023–7.
49. Lee HH, Chernesky MA, Schachter J, Burczak JD, Andrews WW, Muldoon S, et al. Diagnosis of Chlamydia trachomatis genitourinary infection in women by ligase chain reaction assay of urine. *Lancet*. 1995;345(8944):213–6.
50. Nagot N, Meda N, Ouangre A, Ouedraogo A, Yaro S, Sombie I, et al. Review of STI and HIV epidemiological data from 1990 to 2001 in urban Burkina Faso: implications for STI and HIV control. *Sex Transm Infect*. 2004;80(2):124–9.
51. World Health Organization. Risk Factors [Internet]. 2014. Available from: [http://www.who.int/topics/risk\\_factors/en/](http://www.who.int/topics/risk_factors/en/)
52. Adams EJ, Charlett A, Edmunds WJ, Hughes G. Chlamydia trachomatis in the United Kingdom: a systematic review and analysis of prevalence studies. *Sex Transm Infect*. 2004;80(5):354–62.
53. Radcliffe KW, Ahmad S, Gilleran G, Ross JD. Demographic and behavioural profile of adults infected with chlamydia: a case-control study. *Sex Transm Infect*. 2001;77(4):265–70.
54. Han Y, Coles FB, Hipp S. Screening criteria for Chlamydia trachomatis in family planning clinics: accounting for prevalence and clients' characteristics. *Fam Plann Perspect*. 1997;29(4):163–6.
55. Hart G. Factors associated with genital chlamydial and gonococcal infection in females. *Genitourin Med*. 1992;68(4):217–20.
56. Miranda AE, Szwarcwald CL, Peres RL, Page-Shafer K. Prevalence and risk behaviors for chlamydial infection in a population-based study of female adolescents in Brazil. *Sex Transm Dis*. 2004;31(9):542–6.

57. LaMontagne DS, Fenton KA, Randall S, Anderson S, Carter P. Establishing the National Chlamydia Screening Programme in England: results from the first full year of screening. *Sex Transm Infect.* 2004;80(5):335–41.
58. Imai H, Nakao H, Shinohara H, Fujii Y, Tsukino H, Hamasuna R, et al. Population-based study of asymptomatic infection with *Chlamydia trachomatis* among female and male students. *Int J STD AIDS.* 2010;21(5):362–6.
59. Marrazzo JM, Handsfield HH, Whittington WLH. Predicting chlamydial and gonococcal cervical infection: implications for management of cervicitis. *Obstet Gynecol.* 2002;100(3):579–84.
60. Hilger TM, Smith EM, Ault K. Predictors of *Chlamydia trachomatis* infection among women attending rural Midwest family planning clinics. *Infect Dis Obstet Gynecol.* 2001;9(1):3–8.
61. Gershman KA, Barrow JC. A tale of two sexually transmitted diseases. Prevalences and predictors of chlamydia and gonorrhoea in women attending Colorado family planning clinics. *Sex Transm Dis.* 1996;23(6):481–8.
62. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol.* 2008;61(11):1085–94.
63. Moons KGM, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol.* 2002;55(7):633–6.
64. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol.* 1996;49(12):1373–9.
65. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol.* 2007;165(6):710–8.
66. Hosmer DW, Lemeshow S. *Applied logistic regression.* Second Edi. Statistics in Medicine. New York: John Wiley & Sons; 2000.
67. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* First edit. Springer; 2002.
68. Little RJ, Rubin D. *Statistical Analysis with Missing Data.* New York: John Wiley and Sons; 1987.
69. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* 2006;59(10):1092–101.
70. Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Stat Methods Med Res.* 2007;16(3):277–98.
71. Vergouwe Y, Royston P, Moons KGM, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol.* 2010;63(2):205–14.
72. Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol.* 2010;63(7):721–7.
73. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 2000;19(4):453–73.
74. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001;54(8):774–81.
75. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361–87.
76. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128–38.
77. Miller ME, Langefeld CD, Tierney WM, Hui SL, McDonald CJ. Validation of probabilistic predictions. *Med Decis Making.* 1993;13(1):49–58.
78. Hilden J. The area under the ROC curve and its competitors. *Med Decis Making.* 1991;11(2):95–101.

79. 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.* 2008;61(1):76–86.
80. Holland JL, Wilczynski NL, Haynes RB. Optimal search strategies for identifying sound clinical prediction studies in EMBASE. *BMC Med Inform Decis Mak.* 2005;5:11.
81. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews. *PLoS One.* 2012;7(2):e32844.
82. Ingui BJ, Rogers MAM. Searching for Clinical Prediction Rules in Medline. *J Am Med Informatics Assoc. American Medical Informatics Association;* 2001;8(4):391–7.
83. Moons KGM, Donders ART, Steyerberg EW, Harrell FE. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. *J Clin Epidemiol.* 2004;57(12):1262–70.
84. Cessie S Le, Houwelingen JC Van. Ridge Estimators in Logistic Regression. *Appl Stat. JSTOR;* 1992;41(1):191–201.
85. Su YS, Gelman A, Hill J, Yajima M. Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box. *J Stat Softw.* 2011;45(2):1–31.
86. StataCorp. *Stata Statistical Software.* College Station, TX: StataCorp LP; 2011.
87. R Development Core Team. *R: A language and environment for statistical computing, reference index [Internet].* Vienna, Austria: R Foundation for Statistical Computing; 2005. Available from: <http://www.r-project.org>.
88. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* 2006;59(10):1092–101.
89. Sellors JW, Pickard L, Gafni A, Goldsmith CH, Jang D, Mahony JB, et al. Effectiveness and efficiency of selective vs universal screening for chlamydial infection in sexually active young women. *Arch Intern Med.* 1992;152(9):1837–44.
90. Al-Tayyib AA, Miller WC, Rogers SM, Leone PA, Law DCG, Ford CA, et al. Evaluation of risk score algorithms for detection of chlamydial and gonococcal infections in an emergency department setting. *Acad Emerg Med.* 2008;15(2):126–35.
91. Van Valkengoed IG, Boeke AJ, Morr e SA, van den Brule AJ, Meijer CJ, Devill e W, et al. Disappointing performance of literature-derived selective screening criteria for asymptomatic *Chlamydia trachomatis* infection in an inner-city population. *Sex Transm Dis.* 2000;27(9):504–7.
92. Stein CR, Kaufman JS, Ford CA, Leone PA, Feldblum PJ, Miller WC. Screening young adults for prevalent chlamydial infection in community settings. *Ann Epidemiol.* 2008;18(7):560–71.
93. Van Duynhoven YT, van de Laar MJ, Fennema JS, van Doornum GJ, van den Hoek JA. Development and evaluation of screening strategies for *Chlamydia trachomatis* infections in an STD clinic. *GenitourinMed.* 1995;71(6):375–81.
94. Johnson BA, Poses RM, Fortner CA, Meier FA, Dalton HP. Derivation and validation of a clinical diagnostic model for chlamydial cervical infection in university women. *JAMA.* 1990;264(24):3161–5.
95. 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.* 2005;58(5):475–83.
96. Shields SA, Wong T, Mann J, Jolly AM, Haase D, Mahaffey S, et al. Prevalence and correlates of *Chlamydia* infection in Canadian street youth. *J Adolesc Health.* 2004;34(5):384–90.
97. Wang Q, Yang P, Zhong M, Wang G. Validation of diagnostic algorithms for syndromic management of sexually transmitted diseases. *Chin Med J (Engl).* 2003;116(2):181–6.
98. Verhoeven V, Avonts D, Meheus A, Goossens H, Ieven M, Chapelle S, et al. Chlamydial infection: an accurate model for opportunistic screening in general practice. *Sex Transm Infect.* 2003;79(4):313–7.
99. Passey M, Mgone CS, Lupiwa S, Tiwara S, Lupiwa T, Alpers MP. Screening for sexually transmitted diseases in rural women in Papua New Guinea: are WHO therapeutic algorithms appropriate for case detection? *Bull World Health Organ.* 1998;76(4):401–11.

100. La montagne DS, Patrick LE, Fine DN, Marrazzo JM. Re-evaluating Selective Screening Criteria for Chlamydial Infection Among Women in the U.S. Pacific Northwest. *Sex Transm Dis.* 2004;31(5):283–9.
101. Van Valkengoed IG, Morré SA, van den Brule AJ, Meijer CJ, Devillé W, Bouter LM, et al. Low diagnostic accuracy of selective screening criteria for asymptomatic *Chlamydia trachomatis* infections in the general population. *Sex Transm Infect.* 2000;76(5):375–80.
102. Alary M, Baganizi E, Guèdèmè A, Padonou F, Davo N, Adjovi C, et al. Evaluation of clinical algorithms for the diagnosis of gonococcal and chlamydial infections among men with urethral discharge or dysuria and women with vaginal discharge in Benin. *Sex Transm Infect.* 1998;74 Suppl 1:S44–9.
103. Fonck K, Kidula N, Jaoko W, Estambale B, Claeys P, Ndinya-Achola J, et al. Validity of the vaginal discharge algorithm among pregnant and non-pregnant women in Nairobi, Kenya. *Sex Transm Infect.* 2000;76(1):33–8.
104. Guimarães EMB, Guimarães MDC, Vieira MAS, Bontempo NM, Seixas MSS, Garcia MSD, et al. Lack of utility of risk score and gynecological examination for screening for sexually transmitted infections in sexually active adolescents. *BMC Med.* 2009;7:8.
105. Latino MA, Caneparo A, Rosso C, De MD, De IG, Intorcchia P, et al. Prevalence and risk factors for *Chlamydia trachomatis* infection in young women in north-west of Italy. *Minerva Ginecol.* 2008;60(1):29–37.
106. Massé R, Laperrière H, Rousseau H, Lefebvre J, Remis RS. *Chlamydia trachomatis* cervical infection: prevalence and determinants among women presenting for routine gynecologic examination. *CMAJ.* 1991;145(8):953–61.
107. Miller WC, Hoffman IF, Owen-O'Dowd J, McPherson JT, Privette A, Schmitz JL, et al. Selective screening for chlamydial infection: which criteria to use? *Am J Prev Med.* 2000;18(2):115–22.
108. Reed JL, Mahabee-Gittens EM, Huppert JS. A decision rule to identify adolescent females with cervical infections. *J Womens Health (Larchmt).* 2007;16(2):272–80.
109. Ryan CA, Courtois BN, Hawes SE, Stevens CE, Eschenbach DA, Holmes KK. Risk assessment, symptoms, and signs as predictors of vulvovaginal and cervical infections in an urban US STD clinic: implications for use of STD algorithms. *Sex Transm Infect.* 1998;74 Suppl 1:S59–76.
110. Verhoeven V, Avonts D, Van Royen P, Weyler J, Wang X, Stalpaert M. Performance of a screening algorithm for chlamydial infection in 2 samples of patients in general practice. *Scand J Infect Dis.* 2004;36(11-12):873–5.
111. Hocking J, Fairley CK. Do the characteristics of sexual health centre clients predict chlamydia infection sufficiently strongly to allow selective screening? *Sex Health.* 2005;2(3):185–92.
112. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009;338(b605).
113. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. *Stat Med.* 2002;21(24):3803–22.
114. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol.* 1996;49(8):907–16.
115. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med.* 2006;25(1):127–41.
116. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med.* 2003;29(7):1043–51.
117. 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.* 2010;172(8):971–80.
118. Ting KM, Witten IH. Issues in Stacked Generalization. *J Artif Intell Res.* 2011 May 26;10(1):271–89.
119. Clarke B. Comparing Bayes Model Averaging and Stacking When Model Approximation Error Cannot Be Ignored. *J Mach Learn Res.* 2003;4:683–712.

120. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med.* 2006;144(3):201–9.
121. Sloan NL, Winikoff B, Haberland N, Coggins C, Elias C. Screening and syndromic approaches to identify gonorrhea and chlamydial infection among women. *Stud Fam Plann.* 2000 Mar;31(1):55–68.