

INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ITBA

ESCUELA DE INGENIERÍA Y TECNOLOGÍA

SISTEMA DE SOPORTE A LA DECISIÓN CLÍNICA PARA EVALUACIÓN CUANTITATIVA DE RIESGO DE CÁNCER DE MAMA

AUTOR: Tajerian, Matías Nazareth (Leg. N° 53332)

PROFESOR RESPONSABLE O TUTOR/ES: Dr. Carlos Otero y Dra. Luciana Rubin

TRABAJO FINAL PRESENTADO PARA LA OBTENCIÓN DEL TÍTULO DE BIOINGENIERO

BUENOS AIRES

SEGUNDO CUATRIMESTRE, 2017

*Dedicado a
todos los que escribieron
algún capítulo en mi vida.
En especial a mi viejo.*

”Los progresos de la medicina y de la bioingeniería podrán considerarse verdaderos logros para la humanidad cuando todas las personas tengan acceso a sus beneficios y dejen de ser un privilegio para las minorías”.

René Favalaro

Buenos Aires, Congreso de Bioingeniería, 1999

Agradecimientos

El contexto y las circunstancias nos construyen día a día, y cada logro es inconcebible sin ese viento de la vida. Agradezco a los afectos, a las enseñanzas, a las compañías y a los hogares que fueron, son y serán ese viento.

Resumen

El cáncer es la segunda causa de muerte en el mundo, siendo responsable de 1 de cada 6 muertes. En particular, el cáncer de mama es el más frecuente entre las mujeres y es el de mayor incidencia y mortalidad en Argentina. Es sabido que una detección temprana es crucial para el tratamiento de la enfermedad. En este sentido, la evaluación de riesgo es una herramienta de gran utilidad.

Este proyecto se secciona en dos partes. En la primera se estudian cuatro modelos de evaluación de riesgo de cáncer de mama: los modelos de Gail, Claus, BRCAPRO y Tyrer-Cuzick. Se analizan sus métodos, factores de riesgo, limitaciones, entre otros aspectos. Finalmente, se procede a seleccionar uno de ellos para su implementación. La segunda parte consta del desarrollo de un sistema de soporte a la decisión clínica para evaluación cuantitativa de riesgo. En ésta se incorpora el modelo seleccionado, y se lleva a cabo un software para realizar el cálculo de riesgo, a partir del método en cascada.

El trabajo realizado bajo la dirección de profesionales del Hospital Italiano de Buenos Aires, tiene una concreta proyección para ser aplicado en la historia clínica electrónica, con el objetivo de asistir a los profesionales de salud de atención primaria para la detección de pacientes de alto riesgo.

Índice general

Agradecimientos	III
Resumen	IV
Índice general	V
Lista de figuras	VII
Lista de tablas	IX

I Análisis y evaluación de modelos estadísticos para cálculo de riesgo de cáncer de mama (CM) 1

1. Introducción	2
1.1. Motivación	2
1.2. Sistemas de soporte a la decisión clínica (CDSS)	3
1.3. Factores de riesgo de CM	5
1.3.1. BRCA I y BRCA II	7
1.4. Importancia de evaluación de riesgo de CM	7
2. Modelos de Evaluación de Riesgo de Cáncer de Mama	9
2.1. Modelo de Gail	11
2.1.1. Modelo de Gail en la población argentina	21
2.2. Modelo de Claus	22
2.3. BRCAPRO	26
2.4. Modelo Tyrer-Cuzick o IBIS	31
3. Selección de modelo de evaluación	35

II	Diseño y desarrollo de CDSS basado en modelo de Gail	42
4.	Etapas de vida del CDSS	43
4.1.	Requerimientos	44
4.2.	Arquitectura y diseño	45
4.2.1.	Servicio web	46
4.2.2.	Diseño de programa	47
4.2.3.	Interoperabilidad y estándares HL7	49
4.3.	Implementación	50
4.3.1.	Módulo comunicación	50
4.3.2.	Interfaz de usuario	51
4.3.3.	Módulo procesamiento	52
4.4.	Verificación del algoritmo	53
4.5.	Validación y puesta en marcha	53
5.	Conclusión y desafíos a futuro	55
A.	Ejemplo de estándar FHIR en JSON para evaluación de riesgo	57
B.	Módulo de comunicación	59
C.	Método para cálculo de riesgo	61
D.	Método para cálculo de probabilidad	66
	Bibliografía	67

Índice de figuras

1.1. Arquitectura de CDSS basados en conocimientos [1]	4
1.2. Teorema fundamental de la informática biomédica de Friedman [2]	4
1.3. Estadística de la National Breast Cancer Foundation. Entre el 60 % y 70 % de los pacientes con CM no se conocen los factores de riesgo [3]	6
1.4. Árbol de decisión. Se considera alto riesgo cuando la probabilidad de portar el gen mutado, o desarrollar CM en un período de 5 años son $\ast > 10\%$, $\dagger > 1,67\%$, respectivamente [4]	8
2.1. Probabilidad acumulada de CM para mujer con dos parientes de segundo grado (materno y paterno) por edad de detección [5]	24
2.2. Caso: 25 % y 33 % son los valores de riesgo a lo largo de la vida de desarrollar CM, según modelo de Claus y según su correspondiente modificatoria [6], respectivamente.	25
2.3. Riesgo acumulado para portadores (A) y no portadores de mutación (C), incidencia específica por edad para ambos grupos (B y D), y likelihood ratio y probabilidad a posteriori de portar la mutación en paciente con enfermedad desarrollada (E) y paciente sin enfermedad desarrollada (F). Tanto para CM como para CO [7].	29
2.4. Tasas acumuladas para CM y CO en portadores de BRCA I (arriba) y BRCA II (abajo) [8]	30
2.5. Proceso de cálculo de riesgo en pacientes según el modelo IBIS [9].	34
3.1. Curvas ROC para los 4 modelos [10]	36
3.2. Valores de AUC para distintas validaciones del modelo de Gail [11]	38
3.3. Valores de E/O para distintas validaciones del modelo de Gail [11]	39

3.4. Árbol de decisión para elección de modelo de evaluación de riesgo de CM. [4]	41
4.1. Ciclo de vida de software. Modelo de cascada realimentado.	44
4.2. Arquitectura de CDSS de evaluación de riesgo de CM	45
4.3. Arquitectura de servicio web	47
4.4. Flujograma del proceso de soporte a la decisión	48
4.5. Módulos del CDSS	49
4.6. Mensaje de advertencia	51
4.7. Interfaz de usuario	52

Índice de tablas

2.1.	Tabla con riesgos relativos asociados a cada factor de riesgo contemplado en el modelo (normalizados) [12]	14
2.2.	Tasas de incidencia específica por edad, h_1^* . [13]. Modelo Gail 1 con tasas para todos los casos de CM (BCDDP). Modelo Gail 2 únicamente CM invasivo (SEER).	15
2.3.	Proyección de probabilidades de desarrollar CM dentro de 10, 20 y 30 años [12]	16
2.4.	Estimación de riesgos relativos asociados a factores de riesgo según diferentes estudios [13].	18
2.5.	Cantidad E, O y cociente E/O para el modelo original y el modificado, a partir de la muestra placebo del Breast Cancer Prevention Trial [13]	20
2.6.	Resultados de evaluación por máxima verosimilitud para diferentes modelos de segregación [14]	22
2.7.	Tasas entre riesgo de CM en heterocigotas y riesgo de CM en homocigotas [14]	23
2.8.	Probabilidad acumulada para mujer con uno o dos parientes de primer grado con CO [15]	25
2.9.	Los seis fenotipos con sus respectivas expresiones de riesgo [9]	32
2.10.	Estimaciones de riesgo acumulado para portadores de mutaciones [9]	32
2.11.	Riesgos relativos del modelo de IBIS [9]	33
3.1.	Áreas bajo la curva ROC para modelos de Gail, Claus, BRCA-PRO y Tyrer-Cuzick según estudio de Amir [10].	36
3.2.	Cantidad de casos esperados vs. observados para los 4 modelos ($n = 3150$) [10]	36
3.3.	Variables consideradas en distintos modelos de evaluación de riesgo [16]	40
4.1.	Riesgos relativos asociado a detección de hiperplasias atípicas (AH) [17]	45

Parte I

Análisis y evaluación de modelos estadísticos para cálculo de riesgo de cáncer de mama (CM)

Capítulo 1

Introducción

El presente Proyecto Final de Carrera se llevó a cabo en conjunto entre el Departamento de Bioingeniería del Instituto Tecnológico de Buenos Aires (ITBA) y el Departamento de Informática en Salud del Hospital Italiano de Buenos Aires (HIBA). Para este trabajo se plantean dos objetivos. El primero es realizar la investigación y el análisis de los diferentes modelos de riesgo para cálculo de cáncer de mama (CM), y el segundo es el desarrollo de un sistema de soporte de decisión clínica (CDSS) para evaluación cuantitativa de riesgo de CM que permita la identificación de pacientes con alto riesgo en la población.

1.1. Motivación

Según la Organización Mundial de la Salud, el cáncer es la segunda causa de muerte en el mundo, siendo responsable de 1 de cada 6 muertes. Y para los próximos 20 años, se prevé un aumento del 70% en nuevos casos de esta enfermedad. Sólo una cuarta parte de las muertes causadas por cáncer sucede en países de altos ingresos y esto se debe a las posibilidades de ofrecer tratamientos a enfermos oncológicos y al fomento de políticas de lucha contra el cáncer [18].

En cuanto al CM, es el más común en las mujeres. En el mundo, es la segunda mayor causa de muerte dentro de las muertes causadas por cáncer. En Argentina, es la primera, provocando 5600 muertes por año, y es el cáncer de mayor incidencia. Datos del Ministerio de Salud de la Nación reflejan la desigualdad en la tasa de mortalidad (Jujuy y Santiago del Estero las menores, mientras que San Luis y CABA son las mayores). En otras palabras, la Ciudad de Buenos Aires es a nivel regional un foco con altas tasas relativas

de incidencia de CM [19].

Es sabido que la detección temprana, no sólo del CM, sino de cualquier tipo de cáncer, es fundamental para un tratamiento eficaz y para aumentar la tasa de supervivencia. Por ello es que toda herramienta de aplicación clínica que contribuya en esta dirección puede ser de suma relevancia. Entre dichas herramientas clínicas se encuentran los Sistemas de Soporte a la Decisión Clínica (CDSS).

1.2. Sistemas de soporte a la decisión clínica (CDSS)

Los sistemas de soporte a la decisión clínica son programas informáticos que asisten a los profesionales de la salud en la toma de decisiones con el objetivo de mejorar sus resultados.

Ante un determinado problema, al que le corresponde un proceso de toma de decisión estructurado, estos CDSS utilizan información del paciente, el conocimiento científico y clínico disponible y un motor de reglas para mejorar el resultado final del proceso de toma de decisión. Éstos CDSS son los más comunes y se los denominan basados en conocimientos (knowledge-based), donde el modelo implementado en el sistema deriva del estado del arte respecto a la problemática a tratar. En la Figura 1.1 se presenta un diagrama de la estructura de los CDSS de esta clase. Los datos de entrada son los que se ingresan al sistema y pueden hacer referencia a síntomas, características de la persona, resultados de estudios entre otra información clínica. El motor de inferencia relaciona dichos datos de entrada con el conocimiento en el cual se basa el sistema a partir de reglas lógicas o estadísticas estructuradas que permiten su implementación en un programa, como por ejemplo, un sencillo árbol de decisión condicional, hasta teoremas de cálculo estadístico como el de Bayes. Finalmente, los datos de salida son los resultados del procesamiento en el sistema y deben ser coherentes con el objetivo primario del CDSS para el cual fue desarrollado. Pueden ser valores cualitativos, cuantitativos, posibles diagnósticos, probabilidades, etc. Otro conjunto, menos utilizado, de CDSS son los no basados en el conocimiento. En estos casos, no hay un modelo sustentado por el conocimiento científico a partir del cual se establezcan los algoritmos de inferencia. Para éstos se utilizan métodos con redes neuronales, algoritmos genéticos, machine learning, entre otros.

Estos sistemas pueden aplicarse en diferentes áreas clínicas, como por

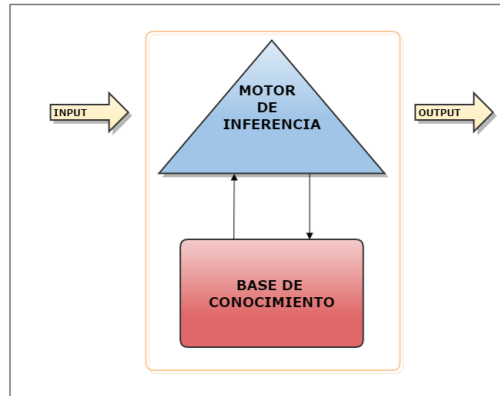


Figura 1.1: Arquitectura de CDSS basados en conocimientos [1]

ejemplo, prescripciones de medicamentos para evitar interacciones entre drogas o reacciones alérgicas, asistencia al diagnóstico, planeamiento de terapias, evaluación de riesgo, entre otras.

El teorema fundamental de la informática biomédica establece que una persona trabajando con un recurso informático es mejor que la misma persona sin dicha asistencia [2]. La Figura 1.2 ilustra esta proposición. Partiendo de este teorema es que se hipotetiza que la implementación del CDSS afectará positivamente en el proceso y en los resultados. Por otra parte, existe evidencia sólida de que los CDSSs aportan en muchos casos significativamente a disminuir la morbilidad en los pacientes, a mejorar la atención primaria, a cumplimentar con medidas preventivas, entre otras áreas de impacto [20,21].



Figura 1.2: Teorema fundamental de la informática biomédica de Friedman [2]

En particular, para el presente proyecto, el CDSS a desarrollar pertenece al área de evaluación de riesgo, entendiéndose como el proceso por el cual se identifican y ponderan los factores que contribuyen a la probabilidad de ocurrencia del daño. En la evaluación de riesgo de CM, se toma información de los factores de riesgo del paciente, y en base a modelos matemáticos y estadísticos embebidos en el programa, se estima la probabilidad de desarrollar

la enfermedad, para que el profesional médico pueda proceder de la mejor manera posible [22–24].

A modo de anticipo, las salidas en los modelos que se analizarán en este trabajo pueden ser, por un lado, la probabilidad de desarrollar CM, y por otro, la probabilidad de ser portador del gen BRCA I o II mutado.

1.3. Factores de riesgo de CM

El CM es una enfermedad que se caracteriza por la proliferación anómala y descontrolada de las células del tejido mamario, que afectan el tejido circundante, pudiendo propagarse por el resto del cuerpo generando metástasis.

Hoy en día no es posible identificar la causa que genera la enfermedad, pero sí se han identificado varios factores de riesgo, es decir, hábitos personales o exposiciones ambientales que aumentan la probabilidad de desarrollar la enfermedad (Figura 1.3) [25]. La National Breast Cancer Foundation lista los siguiente factores [3]:

- Género: La enfermedad es 100 veces más frecuente en mujeres que en hombres.
- Edad: A partir de los 55 años, el riesgo de contraer la enfermedad aumenta. De hecho, por definición, se considera a la mujer de avanzada edad como paciente de alto riesgo.
- Etnia: Diferentes etnias tienen distintos grados de incidencia de la enfermedad. Así, en mujeres caucásicas el riesgo es mayor que entre las mujeres africanas.
- Historia familiar: El hecho de poseer algún familiar, (madre, hermana, tía) que haya contraído CM aumenta el riesgo.
- Historia personal de salud: La detección de lesiones precancerosas, hiperplasias, carcinomas in situ ductal o lobulillar, son otros factores de riesgo. También la detección de cáncer en un pecho, aumenta las posibilidades de contraer la enfermedad en el otro.
- Historia de período menstrual y reproductivo: Las edades tempranas y tardes de menarca y menopausia, respectivamente, como ser madre a una edad tardía, son factores que aumentan el riesgo de CM.

- Cambios genómicos: Se han descubierto que mutaciones en el gen BRCA I y BRCA II son factores de alto riesgo que pueden estar vinculados con la historia familiar, y pueden ser transmitidos a la descendencia.
- Tejido mamario denso: No sólo es un factor de riesgo el hecho de tener tejido mamario denso, sino que también dificulta la detección de lesiones cancerosas en mamografías.

Existen otros factores que aumentan el riesgo, pero que pueden ser modificados, como por ejemplo, el estilo de vida sedentario, una dieta alta en grasas saturadas, el sobrepeso, consumo de alcohol, la exposición a la radiación, y las terapias de reemplazo hormonal (HRT).

Los modelos matemáticos, toman algunos de estos factores para llevar a cabo el cálculo de riesgo en los pacientes. La elección de los factores determina en muchos casos, la utilidad y el fin del modelo.

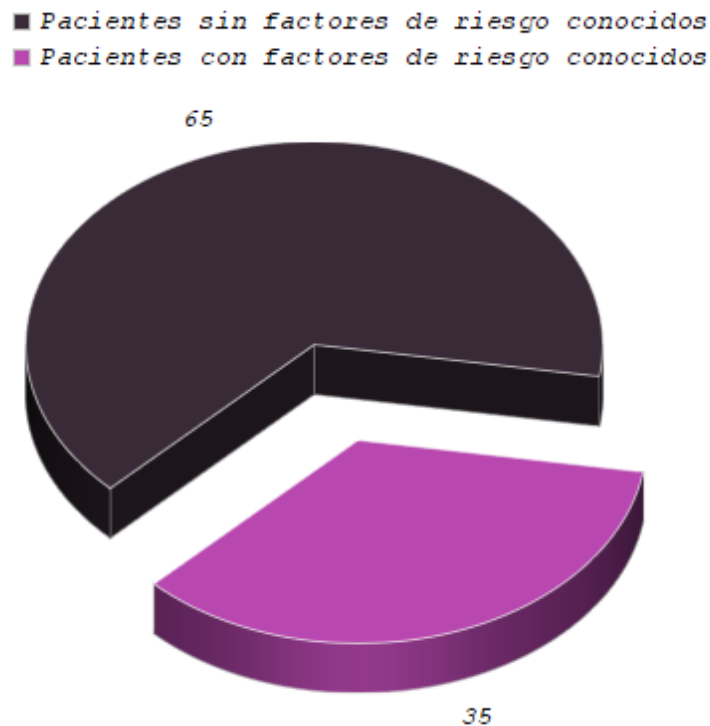


Figura 1.3: Estadística de la National Breast Cancer Foundation. Entre el 60% y 70% de los pacientes con CM no se conocen los factores de riesgo [3]

1.3.1. BRCA I y BRCA II

Los genes BRCA I y BRCA II producen proteínas que contribuyen a la reparación de ADN dañado. Cuando estos genes sufren alguna mutación, y dichas proteínas pierden funcionalidad o el gen deja de expresarse correctamente, la probabilidad de que surjan células tumorales aumenta sensiblemente. Por ello, las mutaciones en BRCA I y II están fuertemente vinculados al desarrollo de CM y CO. Si bien dichas mutaciones son las de mayor penetrancia en CM, es decir que la proporción de los individuos que expresan el fenotipo patológico es alta, existen otras de menor penetrancia, menos estudiadas, que pueden causar la enfermedad como por ejemplo anomalías en STK11, ATM, TP53, RAD51C, entre otros [26].

Hoy en día, con estudios genéticos es posible identificar las mutaciones en los pacientes. Si bien los costos de estos estudios han descendido notablemente en los últimos años, actualmente no dejan de ser caros, entre 300 y 5000 dólares dependiendo la cantidad de áreas y genes que se examinen. En parte, el objetivo de cálculo de riesgo del BRCA mutado es la correcta identificación de los potenciales pacientes para la realización de estos estudios.

1.4. Importancia de evaluación de riesgo de CM

Si bien las tasas de mortalidad en los últimos años han bajado, entre otras causas, por el cese de tratamientos de reemplazo hormonal, las mejoras en los tratamientos de CM y en técnicas de diagnóstico como las mamografías digitales, la incidencia del CM en muchas regiones está en aumento y se debe a cambios de hábitos reproductivos, sedentarismo, aumento de índices de obesidad [27, 28].

Consecuentemente, por un lado sabiendo que la detección temprana es crucial para un tratamiento efectivo, y considerando por otro lado que en los casos de genotipos patológicos no hay manifestación fenotípica más allá del desarrollo de la enfermedad que pueda identificar a la paciente de alto riesgo, es de gran utilidad la evaluación de riesgo en CM.

Una eficaz evaluación de riesgo permite definir las medidas preventivas o intervenciones profilácticas óptimas para cada perfil. En la Figura 1.4 se muestra un árbol de decisión generalizado que vincula algunas de las medidas preventivas, como pueden ser los cambios de hábitos, la vigilancia periódica o la mastectomía para pacientes de alto riesgo, con los resultados de las dos clases de evaluación de riesgo [4].

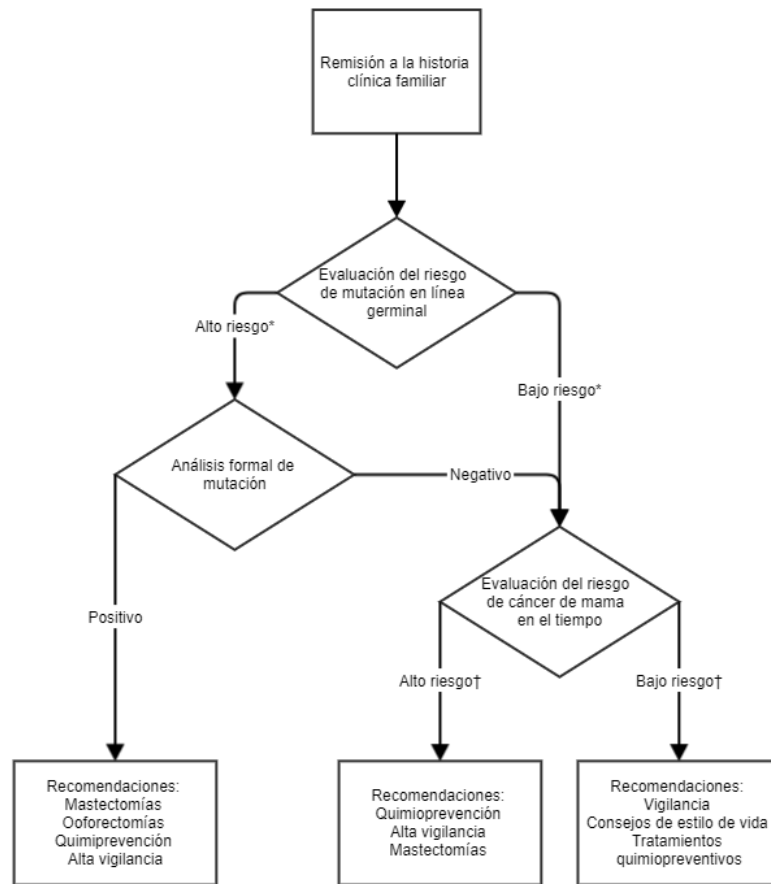


Figura 1.4: Árbol de decisión. Se considera alto riesgo cuando la probabilidad de portar el gen mutado, o desarrollar CM en un período de 5 años son $*$ $> 10\%$, $\dagger > 1,67\%$, respectivamente [4]

En los Capítulos 2 y 3, se presentará con mayor detalle los fines para los que pueden utilizarse los diferentes modelos, evaluando sus ventajas y limitaciones.

Capítulo 2

Modelos de Evaluación de Riesgo de Cáncer de Mama

Los modelos matemáticos para el cálculo de riesgo de CM, son hoy en día una herramienta de gran utilidad para calcular la probabilidad de desarrollar la enfermedad, y en función de eso generar cambios de hábitos, adoptar un seguimiento intensivo de los pacientes, incluso, en pacientes de alto riesgo, prescribir tratamientos quimiopreventivos o proceder a mastectomías [24].

Existen varios tipos de modelos, y cada uno posee ventajas y desventajas propias, contemplan un número de variables determinado y dejan de lado otra cantidad de información, dependiendo del foco y del objetivo del modelo.

No existe actualmente uno que permita abarcar todas las situaciones posibles ni todos los criterios clínicos a la hora del cálculo de riesgo. Es por eso que dependiendo el contexto del paciente, un modelo puede ser más o menos preciso. Como cada caso particular puede precisar de modelos distintos para realizar una evaluación óptima, es necesario conocer los ejes principales de cada uno, así como las situaciones que no contemplan, para saber cuál o cuáles aplicar en los diferentes casos, y así obtener una evaluación de riesgos lo más acertada posible.

Hay dos grandes grupos de modelos para evaluación de riesgo. Por un lado, los que estiman la probabilidad de desarrollar CM. En este conjunto están aquellos que analizan los factores de riesgo ambientales y/o personales, como el modelo de Gail, uno de los más utilizados para cálculo de riesgo o el modelo de Claus, que evalúa el riesgo en función de la historia familiar.

Por otro lado, hay un conjunto de modelos que permiten estimar la probabilidad de mutación de los genes BRCA I o BRCA II. Estas mutaciones conllevan un alto riesgo asociado al CM. En este grupo se encuentra, por ejemplo, el modelo BRACAPRO.

Finalmente, se describe el modelo Tyrer Cuzick que combina tanto cálculo de riesgo de mutación como de desarrollo de la enfermedad.

Es importante destacar que en ninguno de los cuatro casos se ha encontrado entre la bibliografía disponible, estudios o información acerca de procesos de evaluación o validación en la población argentina. Si bien escapa al alcance de este proyecto, se plantea dicho análisis como un importante desafío a futuro.

2.1. Modelo de Gail

El modelo de Gail es uno de los modelos más conocidos y difundidos en el cálculo de riesgo de CM. Fue propuesto originalmente en 1898 [12], convirtiéndose en el primer modelo para predicción de riesgo de CM en mujeres blancas.

La naturaleza multifactorial del CM, llevó a los investigadores a abordar el problema de la estimación de riesgo en tres partes:

- Selección de los factores de riesgo y estimación del riesgo relativo para combinación de dichos factores
- Estimación de la tasa básica de riesgo de CM específica de la edad
- Proyección de probabilidades a largo plazo de desarrollar CM

La resolución del primer problema se llevó a cabo a partir de los datos de pacientes del Breast Cancer Detection Demonstration Project, programa de screening de CM realizado entre los años 1973 y 1980. El programa constó de un estudio de caso-control de seguimiento de 64.182 pacientes (de las 283.222 mujeres dentro del proyecto).

En particular, el análisis para el desarrollo del modelo de Gail se basó en un total de 5998 personas de las cuales 3146 eran controles y 2852 casos.

Se implementó el método de regresión logística para el análisis de los datos, ya que la variable dependiente de esta primera parte del modelo es del tipo categórico (saludable/enfermo). Las variables independientes evaluadas representan diferentes factores de riesgo que se detectaron en la población de estudio. Luego del análisis se determinó que los predictores de mayor impacto en el desarrollo de CM eran:

- Antecedentes de CM en familiares de primer grado (madre, hermanas)
- Edad tardía de primer parto
- Edad temprana de menarca
- Múltiples biopsias benignas de mama

Además, los resultados determinaron que la edad es también uno de los factores principales en el riesgo asociado. Es decir que a mayor edad, la probabilidad de contraer la enfermedad aumenta. Por otra parte, se encontró interrelación entre los factores de cantidad de biopsias y edad de la persona, así como también entre la edad de primer parto y la cantidad de familiares con CM. Estas interacciones fueron plasmadas en el modelo. Es importante

destacar también que en cuanto a los antecedente familiares, el modelo de Gail no contempla el linaje paternal. Tampoco considera si el CM en los parientes es bilateral (desarrollado en ambas mamás) ni la edad en la que fue diagnosticada la enfermedad. Asimismo, se omite el riesgo asociado a la detección de hiperplasias atípicas de seno (ductales o lobulillares) en las biopsias, ya habiendo para ese entonces evidencia suficiente de su incidencia en el desarrollo de CM. Dicho factor no fue tenido en cuenta en el modelo debido a que no se disponía de esta información en los individuos estudiados. Modificaciones subsiguientes que se detallarán más adelante incluyen dicho factor de riesgo.

Las abreviaturas respectivas a los factores de riesgo (utilizadas en el artículo de Gail [12]) fueron NUMREL, AGEFLB, AGEMEN y NBIOPS y para la edad de la persona AGECAAT.

La siguiente ecuación, a la cual se arribó a partir del método de regresión logística, describe la razón de posibilidades (OR) ¹ incluyendo los factores mencionados anteriormente. La ecuación obtenida por Gail fue la siguiente:

$$\begin{aligned} \ln\left(\frac{p}{1-p}\right) = & -0,74948 + 0,09401(AGEMEN) + 0,52926(NBIOPS) + \\ & 0,21863(AGEFLB) + 0,95830(NUMREL) + 0,01081(AGECAAT) - \\ & 0,28804(NBIOPS \times AGECAAT) - 0,19081(AGEFLB \times NUMREL) \end{aligned} \quad (2.1)$$

donde $\ln\left(\frac{p}{1-p}\right)$ es el logaritmo natural de OR; NUMREL, AGEFLB, AGEMEN y NBIOPS las abreviaturas utilizadas en el artículo de Gail [12] para los respectivos factores de riesgo y AGECAAT para la edad de la persona. Éstos parámetros toman valores enteros entre 0 y 1, 0 y 2 o 0 y 3, según corresponda.

El modelo de Gail aproxima el riesgo relativo RR con la razón de probabilidad OR ². Modificando algebraicamente la ecuación del modelo se obtiene:

¹Se entiende por razón de posibilidades a la división entre las posibilidades del evento en expuestos a un factor de riesgo por las posibilidades del evento en no expuestos, lo que permite estimar en qué medida ocurre más el evento en expuestos que en no expuestos [25].

²El cálculo de riesgo absoluto se basa en el RR. En el caso de una enfermedad poco frecuente (por ejemplo, casi todos los tipos de cáncer) la razón de posibilidades (OR) es una buena aproximación al riesgo relativo (RR) [25].

$$\begin{aligned}
OR &= e^{-0,74948+0,09401(AGEMEN)+0,52926(NBIOPS)+0,21863(AGEFLB)+...} \\
&= e^{-0,74948} \times e^{0,09401(AGEMEN)} \times e^{0,52926(NBIOPS)} \times e^{0,21863(AGEFLB)}... \\
&= RR_0 \times RR_{AGEMEN} \times RR_{NBIOPS} \times RR_{AGEFLB}...
\end{aligned} \tag{2.2}$$

Si el RR se normaliza con respecto a una mujer de la misma edad pero sin ningún factor de riesgo asociado, se obtienen los factores de riesgo relativo que se muestran en la Tabla 2.1. Por ejemplo, para el riesgo relativo a la edad de menarca < 12, en una mujer de 57 años es:

$$\begin{aligned}
RR_{AGEMEN<12} &= \frac{e^{-0,74948} \times e^{0,01081} \times e^{2 \times 0,09401}}{e^{-0,74948} \times e^{0,01081}} \\
&= \frac{\cancel{e^{-0,74948}} \times \cancel{e^{0,01081}} \times e^{2 \times 0,09401}}{\cancel{e^{-0,74948}} \times \cancel{e^{0,01081}}} \\
&= 1,207
\end{aligned} \tag{2.3}$$

Entonces, los RR estimados para cada factor permiten calcular el RR total, es decir, de origen multifactorial, simplemente multiplicando los RR correspondientes. Por ejemplo el riesgo asociado de una mujer de 53 años, con edad de menarca de 12 años, sin biopsias realizadas, con su primer parto a los 23 años, y con un familiar de primer grado con CM, le corresponde un valor de RR:

$$RR_{total} = 1,099 \times 1,000 \times 2,681 = 2,946$$

De esta manera queda resuelto la primera parte del modelo en donde se buscaban los factores de riesgo y sus respectivos RR.

El problema planteado en el segundo inciso tiene como objetivo hallar los riesgos basales para una determinada edad, sin que la persona esté afectado por factores riesgo. Para ello, se tomaron intervalos de 5 años, desde los 20 a los 79 años y se estimaron las tasas de incidencia para cada uno de ellos (Figura 2.2). Para el cálculo de la incidencia se tomaron en cuenta los casos que fueron detectados en los años 2 y 3 del programa.

Factor de riesgo		Riesgo relativo asociado
AGEMEN (años)		
> 14 (0)		1.000
12-13(1)		1.099
<12 (2)		1.207
NBIOPS		
Edad < 50 años		
0(0)		1.000
1(1)		1.698
>2(2)		2.882
Edad > 50 años		
0(0)		1.000
1(1)		1.273
>2(2)		1.620
AGEFLB (años)	NUMREL	
< 20(0)	0(0)	1.000
	1(1)	2.607
	>2(2)	6.798
20-24 (1)	0(0)	1.244
	1(1)	2.681
	> 2(2)	5.775
25-29 o sin partos (2)	0(0)	1.548
	1(1)	2.756
	>2(2)	4.907
>30 (3)	0(0)	1.927
	1(1)	2.834
	>2(2)	4.169

Tabla 2.1: Tabla con riesgos relativos asociados a cada factor de riesgo contemplado en el modelo (normalizados) [12]

Para el tercer problema, donde se busca la proyección de la probabilidad de desarrollar CM a largo plazo, el artículo de Gail utiliza un método estándar de cálculo [12], y divide en 13 intervalos de edad Δ_j a la población, al igual que en la Tabla 2.2, incluyendo el intervalo $[0, 20)$. Dada la edad (a), el riesgo relativo a partir de los factores (r), y el intervalo de tiempo que se desea pronosticar (τ), se obtiene la probabilidad de desarrollar CM mediante la expresión 2.4

Edades	Modelo Gail 1	Modelo Gail 2	
		Raza blanca	Raza negra
35–39	114.6	66.1	79.6
40–44	203.7	126.5	137.7
45–49	280.8	186.6	165.4
50–54	320.9	221.1	177.9
55–59	293.8	272.1	224.3
60–64	369.4	334.8	275.0
65–69	356.1	392.3	280.3
70–74	307.8	417.8	309.9
75–79	301.3	443.9	360.2

Tabla 2.2: Tasas de incidencia específica por edad, h_1^* . [13]. Modelo Gail 1 con tasas para todos los casos de CM (BCDDP). Modelo Gail 2 únicamente CM invasivo (SEER).

$$P(a, \tau, r) = \sum_j \frac{h_{1j} r_j}{h_{1j} r_j + h_{2j}} \frac{S_1(\tau_{j-1})}{S_1(a)} \frac{S_2(\tau_{j-1})}{S_2(a)} [1 - e^{-\Delta_j(h_{1j} r_j + h_{2j})}] \quad (2.4)$$

donde h_{1j} es el riesgo basal de muerte por CM en el intervalo de edades j , h_{2j} es el riesgo basal de muerte por cualquier otra causa en dicho intervalo de edades, $S_2(t)$ la probabilidad de supervivencia hasta la edad t ante el riesgo de CM, $S_1(t)$ la probabilidad de supervivencia hasta la edad t ante otros riesgos³ y donde j toma los valores necesarios de manera que se contemple el intervalo $[a; a + \tau]$.

Habiendo resuelto las tres etapas planteadas al inicio del artículo, se muestran en la Tabla 2.3 las proyecciones de riesgo a desarrollar CM para diferentes a , τ y r . Es necesario destacar que en el caso en que $a < 50$ y $a + \tau > 50$, el factor de riesgo AGECAAT toma un nuevo valor por lo que se debe calcular el respectivo RR (RR posterior), siempre considerando que el resto de los factores no se modifican.

³Los respectivos valores de supervivencia $S_i(t)$ se obtienen mediante la integral:

$$S_i(t) = e^{-\int_0^t h_i(x) dx}$$

Edad inicial	Años de proyección	RR posterior	RR inicial					
			1.0	2.0	5.0	10.0	20.0	30.0
20	10		0.0	0.1	0.2	0.5	1.0	1.4
	20		0.5	1.0	2.5	4.9	9.5	14.0
	30		1.7	3.4	8.3	15.9	29.3	40.5
30	10		0.5	0.9	2.3	4.4	8.7	12.8
	20		1.7	3.3	8.1	15.6	28.8	39.9
	30	1.0	3.2	4.8	9.5	16.9	29.9	40.8
		2.0	4.7	6.3	10.9	18.2	30.9	41.7
		5.0	8.9	10.4	14.9	21.8	34.0	44.3
		10.0	15.6	17.1	21.2	27.6	38.8	48.3
		20.0	27.6	28.8	32.3	37.8	47.4	55.5
		30.0	37.7	38.7	41.8	46.4	54.7	61.7
40	10		1.2	2.5	6.1	11.8	22.2	31.3
	20	1.0	2.8	4.0	7.5	13.1	23.4	32.4
		2.0	4.3	5.5	8.9	14.5	24.5	33.4
		5.0	8.6	9.7	13.1	18.3	28.0	36.4
		10.0	15.4	16.4	19.5	24.4	33.3	41.1
		20.0	27.4	28.4	30.9	35.2	42.7	49.5
		30.0	37.7	38.5	40.7	44.3	50.8	56.6
	30	1.0	4.4	5.6	9.1	14.6	24.6	33.5
		2.0	7.4	8.6	11.9	17.3	27.0	35.6
		5.0	15.9	17.0	20.0	24.9	33.7	41.5
		10.0	28.3	29.2	31.8	35.9	43.4	50.0
		20.0	47.5	48.1	50.0	53.1	58.5	63.4
		30.0	61.2	61.6	63.1	65.3	69.3	72.8
50	10		1.6	3.1	7.6	14.6	27.1	37.7
	20		3.2	6.4	15.1	27.9	47.8	61.9
	30		4.4	8.5	19.9	35.5	57.8	71.7
60	10		1.8	3.6	8.6	16.5	30.1	41.5
	20		3.0	5.9	14.0	25.9	44.6	58.2
70	10		1.4	2.7	6.7	12.9	24.1	33.7

Tabla 2.3: Proyección de probabilidades de desarrollar CM dentro de 10, 20 y 30 años [12]

Por ejemplo, para una mujer de 40 años, con un RR de 5, tiene una probabilidad de desarrollar CM en los próximos 10 años de 6,1 %. Con los datos de la tabla, y por interpolación bilineal, se pueden estimar las probabilidades que dependan de RR intermedios.

Esta información corresponde al primer modelo de cálculo para riesgo

de CM, publicado en 1989. Varias limitaciones se plantearon en el artículo, que años posteriores fueron razones para modificaciones de este modelo original. El principal limitante que se presenta cada vez que se realizan estudios estadísticos de esta clase es el hecho de si la muestra es suficientemente representativa como para que el modelo sea válido para el análisis de otras poblaciones. Otros estudios han reformulado el modelo de Gail con otras bases de datos y han comparado los resultados con la evidencia observable [13,29]. La Tabla 2.4 muestra los RR calculados para los factores de riesgo en función de cuatro poblaciones blancas, incluida la muestra inicial del modelo (segunda columna).

Factores de riesgo	RR: BCDDP	RR: CASH Study	RR: NHS	RR (95 % intervalo de confianza): BCPT
AGEMEN (años)				
≥14	1.00	1.00	1.00	1.00 (referencia)
12–13	1.10	1.14	1.09	1.21 (0.98–1.50)
≤ 12	1.21	1.29	1.20	1.47 (0.96–2.26)
NBIOPS				
Edad <50 y				
0	1.00	1.00	1.00	1.00 (referencia)
1	1.70	1.97	1.67	1.11 (0.85–1.45)
≥2	2.88	3.89	N/A	1.23 (0.73–2.09)
Edad ≥50 y				
0	1.00	1.00	1.00	1.00 (referencia)
1	1.27	1.77	1.72	1.34 (1.09–1.65)
≥2	1.62	3.13	N/A	1.80 (1.18–2.73)
NUMREL				
AGEFLB <20				
0	1.00	1.00	1.00	1.00 (referencia)
1	2.61	2.04	1.78	2.13 (1.33–3.39)
≥2	6.80	4.16	3.17	4.52 (1.78–11.49)
AGEFLB 20–24				
0	1.24	1.21	1.16	1.53 (1.09–2.16)
1	2.68	2.40	2.03	2.52 (1.46–4.35)
≥2	5.78	4.75	3.55	4.16 (1.88–9.16)
AGEFLB 25–29 o sin nacimientos vivos				
0	1.55	1.47	1.34	2.35 (1.18–4.67)
1	2.76	2.82	2.31	3.00 (1.55–5.80)
≥2	4.91	5.43	3.98	3.82 (1.84–7.93)
AGEFLB ≥30				
0	1.93	1.77	1.56	3.60 (1.29–10.08)
1	2.83	3.32	2.64	3.56 (1.60–7.90)
≥2	4.17	6.20	4.47	3.51 (1.62–7.61)

Tabla 2.4: Estimación de riesgos relativos asociados a factores de riesgo según diferentes estudios [13].

En el estudio de 1999 [13] se propone un modelo de predicción únicamente

para riesgo de CM invasivo, a diferencia del modelo original que las predicciones incluían tanto cáncer invasivo como cáncer in situ ductal y lobulillar. En base a esta premisa se procesan los datos para el cálculo de los RR y para el desarrollo del modelo.

Posteriormente, se comparan los resultados de los riesgos absolutos para las reparametrizaciones a partir de la relación entre enfermos esperados (E) y observados (O), para un grupo de estudio dado. ⁴ La razón E/O es indicador de la calidad del modelo (Tabla 2.5).

⁴La cantidad de casos esperados (E) para una categoría se define como la sumatoria de las probabilidades de las mujeres en dicha categoría.

Variable	Modelo 1: todos los CM	Modelo 2: cáncer invasivo de mama
AGEMEN (años)		
≥ 14	0.85	1.01
12–13	0.92	1.09
≤ 12	0.71	0.91
NBIOPS		
Edad <50 y		
0	0.87	0.71
1 (sin AH)	0.63	0.67
1 (con AH)	0.42	0.51
≥ 2 (sin AH)	1.33	1.76
≥ 2 (con AH)	2.73	2.99
Edad ≥ 50 y		
0	0.82	1.15
1 (sin AH)	0.79	1.04
1 (con AH)	1.03	1.46
≥ 2 (sin AH)	0.73	1.07
≥ 2 (con AH)	0.78	0.77
NUMREL		
AGEFLB <20		
0	—	—
1	1.63	1.69
≥ 2	1.61	1.44
AGEFLB 20–24		
0	0.55	0.87
1	0.78	0.84
≥ 2	0.84	1.12
AGEFLB 25–29 o sin nacimientos vivos		
0	0.54	0.70
1	0.86	1.08
≥ 2	0.71	0.91
AGEFLB ≥ 30		
0	0.61	1.51
1	0.78	1.00
≥ 2	3.03	2.89

Tabla 2.5: Cantidad E, O y cociente E/O para el modelo original y el modificado, a partir de la muestra placebo del Breast Cancer Prevention Trial [13]

Otros estudios incluyeron otras etnias al modelo ya sea validándolo o modificando sus parámetros, como los RR o las tasas de incidencia específicas de edad [30–33]. El modelo fue adaptado y validado para las etnias africanas, asiáticas. Sin embargo, aún hoy, no hay certezas respecto a su eficacia en personas hispanas.

2.1.1. Modelo de Gail en la población argentina

En el marco del Programa Argentino de Consensos de Enfermedades Oncológicas, en Octubre de 2014 se emitió el documento llamado Alto Riesgo Para Cáncer De Mama Consenso Nacional Inter-Sociedades [34]. El objetivo del documento es elaborar un consenso multidisciplinario, que aúne los criterios propuestos por los profesionales que se encuentran involucrados en la prevención, diagnóstico y tratamiento del Cáncer de Mama.

Diferentes instituciones y asociaciones participaron y colaboraron en el desarrollo del documento. Entre ellas, Asociación Médica Argentina, Instituto Nacional del Cáncer, Sociedad Argentina de Genética, Sociedad Argentina de Patología, Sociedad Argentina de Mastología, Sociedad Argentina de Radiología, Sociedad Argentina de Cancerología, Instituto de Oncología Ángel H. Roffo, Asociación Argentina de Oncología Clínica, Sociedad de Cirugía Plástica de Buenos Aires, Asociación Argentina de Ginecología Oncológica, Sociedad Argentina de Terapia Radiante Oncológica.

Este documento explicita que: “En el caso de nuestra población, la aplicabilidad de estos modelos aún no ha sido validada, y por ello el valor de riesgo que arrojan debería ser tomado con cuidado en el proceso de comunicación médico-paciente.” Esto va de la mano con el hecho de que, si bien el modelo de Gail es uno de los más validados, aún no hay estudios que garanticen su validez respecto a la etnia hispana.

Por otra parte, dentro de la unificación de criterios se establece que “el modelo de Gail no es considerado adecuado para el cálculo de riesgo de cáncer de mama en los siguientes casos:

- Familias con cáncer de ovario (CO)
- Familias con afectados de segundo y tercer grado
- Familias con segregación de la enfermedad por rama paterna
- En casos de CM esporádicos el riesgo está sobrestimado
- En casos de CM hereditario el riesgo está subestimado”

2.2. Modelo de Claus

El modelo de Calus, junto con el de Gail, es uno de los métodos más difundidos. Hace foco en pacientes con historia familiar en desarrollo de CM. La hipótesis que sustenta este modelo es la existencia de un alelo autosómico dominante que aumenta sensiblemente la propensión a contraer la enfermedad [14].

Este modelo se desarrolla inicialmente a partir de datos del Cancer and Steroid Hormone Study llevado a cabo por Centers of Disease Control, con 4730 casos de CM con confirmación histológica y 4688 controles. Se recavó información acerca de la historia familiar de cada uno de los pacientes a partir de entrevistas.

Los factores de riesgo que incluye son:

- Historial de familiares de primer y segundo grado afectados por CM
- Edad de detección del CM en los familiares

A partir de un análisis de segregación ⁵ se llegó a la conclusión que el modelo del alelo autosómico dominante es el que mejor describe los datos estudiados. La evaluación de los diferentes modelos de segregación (mixto, locus principal, esporádico, multifactorial) se realizó a partir del método de máxima verosimilitud, dando los valores de la función de verosimilitud que se muestran en la Tabla 2.6.

Modelo	$-2\ln L + c$
Esporádico ($q = H = 0$)	-46,511.36
Multifactorial ($q = 0$)	
Sin diferencia de generación	-46,716.87
Diferencia de generación	-46,717.38
Locus principal ($H = 0$)	
Recesivo	-46,741.72
Intermedio	-46,807.21
Dominante	-46,807.76
Mixto	-46,807.76

Tabla 2.6: Resultados de evaluación por máxima verosimilitud para diferentes modelos de segregación [14]

A partir de este modelo se concluyó que si bien la proporción de CM desencadenado por el gen dominante es considerablemente menor (frecuencia =

⁵El análisis de segregación para los diferentes modelos se llevó a cabo con el programa POINTER [35]

0.0033) al de CM esporádico, casi todas las portadoras del alelo desarrollarán la enfermedad en algún momento de sus vidas y que por ende, es muy importante identificar a estos pacientes de alto riesgo. En particular, se observó que existe un mayor riesgo en edades tempranas, que en edades avanzadas. En la Tabla 2.7, se muestran las proporciones entre CM en pacientes heterocigotas (alelo mutado) y homocigotas normales (sin factores hereditarios), y se evidencia la alta tasa de incidencia del alelo dominante en edades tempranas en comparación con paciente mayores. Prácticamente todos los casos de CM entre 20 y 29 años tienen origen hereditario, y tan sólo el 2% en pacientes de 80 años o más.

Edades	Tasa de riesgo	
	Modelo función escalón	Modelo distribución normal
20-29	98	76
30-39	50	43
40-49	21	25
50-59	13	14
60-69	6	8
70-79	9	5
+80	2	3

Tabla 2.7: Tasas entre riesgo de CM en heterocigotas y riesgo de CM en homocigotas [14]

El método para cálculo de la probabilidad $R(x)$ de desarrollar CM a la edad x está dado por la siguiente expresión:

$$R(x) = \sum_j \sum_k F(x, \theta_j) G_{jk} \left[\frac{G_k f(x_r, \theta_k)}{\sum_k G_k f(x_r, \theta_k)} \right]$$

donde $F(x, \theta_j)$ es la distribución normal de probabilidades y $f(\cdot)$ su respectiva función densidad, x es la edad de la paciente, θ el vector media y varianza, los subíndices j y k recorren los diferentes genotipos (AA, Aa, aa), G_j la probabilidad de que el paciente sea portador del genotipo j , G_{jk} es la probabilidad de que la paciente porte genotipo j dado el genotipo k en su familiar y x_r la edad en la que la madre o hermana de la paciente fue afectada con CM [5].

A partir de este cálculo de probabilidad se puede estimar el riesgo de que la paciente desarrolle CM en un determinado lapso de la siguiente forma:

$$RC(y|x) = \frac{R(y) - R(x)}{1 - R(x)}$$

Donde x es la edad de la paciente sin la enfermedad, e y la edad a la que se proyecta el riesgo. De esta forma, se confeccionaron las tablas de Claus [5], que combinan diferentes posibilidades de riesgo en función de las dos variables anteriormente mencionadas. A modo de ejemplo, se muestra una de estas tablas (Figura 2.1).

Table 7. Predicted Cumulative Probability of Breast Cancer for a Women Who Has One Maternal and One Paternal Second-Degree Relative Affected With Breast Cancer, By Age of Onset of the Affected Relatives

Age of onset of first relative (yr)											
20-29							30-39				
Age of onset of second relative (yr)											
Age (yr)	20-29	30-39	40-49	50-59	60-69	70-79	30-39	40-49	50-59	60-69	70-79
29	.007	.006	.005	.004	.004	.004	.005	.004	.003	.003	.003
39	.023	.020	.017	.016	.015	.014	.016	.014	.012	.011	.010
49	.057	.050	.044	.040	.038	.037	.042	.036	.032	.030	.029
59	.108	.095	.085	.079	.075	.073	.081	.071	.065	.061	.058
69	.160	.143	.130	.121	.116	.113	.124	.111	.102	.097	.094
79	.199	.179	.164	.155	.149	.145	.179	.144	.134	.128	.124
Age of onset of first relative (yr)											
40-49						50-59		60-69		70-79	
Age of onset of second relative (yr)											
Age (yr)	40-49	50-59	60-69	70-79	50-59	60-69	70-79	60-69	70-79	70-79	
29	.003	.003	.002	.002	.002	.002	.002	.002	.001	.001	
39	.011	.010	.009	.008	.008	.007	.006	.006	.005	.005	
49	.030	.026	.024	.023	.022	.020	.019	.018	.016	.015	
59	.061	.055	.050	.048	.048	.044	.041	.039	.037	.034	
69	.098	.089	.083	.080	.080	.074	.071	.068	.065	.062	
79	.128	.118	.112	.108	.108	.102	.098	.095	.091	.087	

Figura 2.1: Probabilidad acumulada de CM para mujer con dos parientes de segundo grado (materno y paterno) por edad de detección [5]

El modelo de Claus tuvo posteriores modificaciones [15]. En 1993 se demostró la correlación entre el CO y el CM a nivel genético. De la misma forma, se elaboró un modelo estadístico que determinó un 13 % de riesgo en contraer CM con un familiar de primer grado habiendo desarrollado CO y un 30 % en el caso de dos familiares (Tabla 2.8). Otras variaciones y ajustes que se llevaron a cabo sobre el método de Claus contemplan la presencia de más de dos familiares con la enfermedad y clasifican el CM según si es o no bilateral [6]. En el árbol genealógico de la Figura 2.2, se resume dicha modificación con la aplicación del modelo rectificado en un ejemplo.

Los fundamentos mendeleianos del modelo de Claus conformaron la antesala al descubrimiento posterior de los genes BRCA I y BRCA II, cuyas mutaciones se vinculan hoy en día con el desarrollo de CM y CO en la mujer.

Edades	Cantidad de familiares de primer grado afectados con CO	
	1	2
29	.002	.011
39	.010	.037
49	.026	.090
59	.054	.164
69	.088	.236
79	.117	.285
89	.135	.308

Tabla 2.8: Probabilidad acumulada para mujer con uno o dos parientes de primer grado con CO [15]

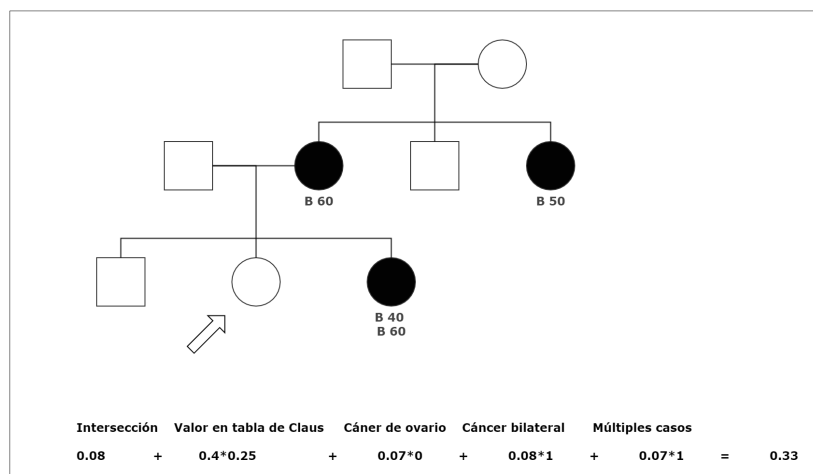


Figura 2.2: Caso: 25 % y 33 % son los valores de riesgo a lo largo de la vida de desarrollar CM, según modelo de Claus y según su correspondiente modificatoria [6], respectivamente.

Si bien el modelo de Claus permite una discriminación estimativa de los pacientes con un alto riesgo de desarrollo de CM, no ha sido validado en muestras poblacionales heterogéneas, de manera tal que se permita su aplicación clínica. Aún así, es una primera herramienta de pre evaluación para potenciales pacientes que necesiten análisis genéticos de mutaciones BRCA.

2.3. BRCAPRO

El modelo BRCAPRO pertenece al segundo conjunto de los mencionados en la introducción de este capítulo. Permite calcular la probabilidad de poseer el gen BRCA I o II mutado, basándose en el método de Bayes, a partir de la historia familiar del paciente.

En BRCAPRO, se contempla:

- Parentesco con todos los familiares de primer y segundo grado
- Edad de inicio de la enfermedad (CM o CO)
- Edad actual o de defunción de los familiares no afectados

El teorema de Bayes tiene una gran relevancia por el hecho de expresar la probabilidad condicional de un suceso A dado B en función de la probabilidad condicional del suceso B dado A. Para este caso, el potencial de esta herramienta se utiliza para calcular la probabilidad de ser portador de la mutación en BRCA, dado una determinada historia familiar. La expresión que rige dicha regla es la siguiente:

$$P(M|H) = \frac{P(H|M)P(M)}{P(H)}$$

donde $P(.)$ son las respectivas probabilidades, M es el individuo con mutación, N su complemento, es decir el individuo sin mutación y H la historia familiar del individuo [7].

Otra forma de expresar el teorema es en función de la razón de verosimilitud (LR), de la siguiente manera:

$$P(M|H) = \frac{LR}{LR + \frac{P(N)}{P(M)}}$$

donde

$$LR = \frac{P(H|M)}{P(H|N)}$$

y donde $\frac{P(N)}{P(M)}$ son las probabilidades a priori de ser portadores de la mutación.

Suponiendo que la probabilidad de un alelo esté mutado es f , entonces $P(M) = 2f(1-f) + f^2 = 2f - f^2$ ya que se considera la probabilidad en cada

uno de los alelos, y los dos simultáneamente. Asimismo, $P(N) = 1 - P(M)$ por ser el complemento.

Las probabilidades para el cálculo de LR se basan en genética mendeleiana. Toda la información familiar se refleja en dicho parámetro. El numerador $P(H|M)$ cuando la información de la familia se ignora en una paciente diagnosticada con CM a los 40 años, es equivalente a la incidencia de la enfermedad a esa edad, entre pacientes con el gen mutado.

Se denota con $B_M(a)$ a la proporción acumulada hasta la edad a de los portadores que fueron diagnosticados con CM. Consecuentemente, $b_M(a)$ es la incidencia específica de la edad, es decir la tasa de cambio en la proporción acumulada [7] ⁶. En los gráficos de las Figuras 2.3 y 2.4 se observan las tasas acumuladas y específicas de CM y CO para BRCA I y las tasas acumuladas para CM y CO de BRCA I y II, respectivamente.

Para este ejemplo, la expresión de LR se la puede formular como:

$$LR = \frac{b_M(40)}{b_N(40)}$$

donde $b_N(40)$ es la incidencia específica de CM a las 40 años entre los no portadores.

Tomando en cuenta las gráficas de la Figura 2.3, $b_M(40) = 0,021$ y $b_N(40) = 0,001$, y el valor $f = 0,0006$ [37], se obtiene que $LR = 21$, $P(M) = 0,0012$, $P(N) = 0,9988$.

$$P(M|H) = \frac{LR}{LR + \frac{P(N)}{P(M)}} = \frac{21}{21 + \frac{0,9988}{0,0012}} = 0,0246$$

Este resultado se interpreta de la siguiente manera: "Dada la historia familiar $H = \text{Individuo con CM detectado a la edad de 40}$, la probabilidad de que sea portadora de mutación en BRCA I es de 2,46 %". Esta probabilidad es 20 veces mayor a la probabilidad a priori $P(M) = 0,12$.

En el caso de $H = \text{Individuo sin CM}$, el LR se calcula con el complemento de las tasas acumuladas, es decir:

$$LR = \frac{1 - B_M(a)}{1 - B_N(a)}$$

⁶Los datos sobre incidencias en mutaciones de BRCA, tanto para CM como para CO, fueron obtenidos de otros estudios [36].

A la hora de considerar en la paciente CM y CO, LR se calcular mediante dos factores correspondientes a cada enfermedad. En caso de que no haya desarrollado ninguna de las dos, (una de las cuatro posibilidades) la expresión sería la siguiente.

$$LR = \frac{1 - B_M(a)}{1 - B_N(a)} \frac{1 - O_M(a)}{1 - O_N(a)}$$

El cálculo de LR se torna más complejo cuando se comienzan a considerar los familiares de primer y segundo grado. El modelo BRCAPRO considera de la historia familiar, como se mencionó anteriormente, las edades de diagnóstico de CM y/o CO, las edades actuales o de fallecimiento de los familiares no afectados, y el parentesco con el individuo. A la hora de calcular el LR, entonces, se contemplan las contribuciones de cada uno de los familiares. Para el numerador, se asume que el individuo es portador de la mutación, se estiman las contribuciones y luego se multiplican los factores de cada familiar. Lo mismo sucede para el denominador, pero en este caso, se supone que el paciente no posee el gen mutado. El ejemplo descrito en el apéndice del estudio *Probability of Carrying a Mutation of Breast–Ovarian Cancer Gene BRCA1 Based on Family History* de Berry [7], muestra con claridad este proceso de cálculo.

Luego del primer trabajo donde se aplicó el teorema de Bayes para el cálculo de probabilidad en el gen BRCA I, Parmigiani [8] publicó una ampliación del modelo para abarcar los genes BRCA I y BRCA II, cuya metodología es adaptable a la incorporación de más genes, ante posibles descubrimientos posteriores.

Si bien fue demostrada su efectividad [38], mejoras subsiguientes fueron perfeccionando el modelo Bayesiano. Por un lado, nuevas etnias o grupos poblacionales con incidencias diferentes fueron incluidos al modelo, como el caso de los judío ashkenazí [39]. Otras modificaciones como la posibilidad de interpretar árboles genealógicos multiétnicos, ajustes por mastectomías en familiares, modelización de penetrancia de CM contralateral, entre otras [40, 41], surgieron con posterioridad.

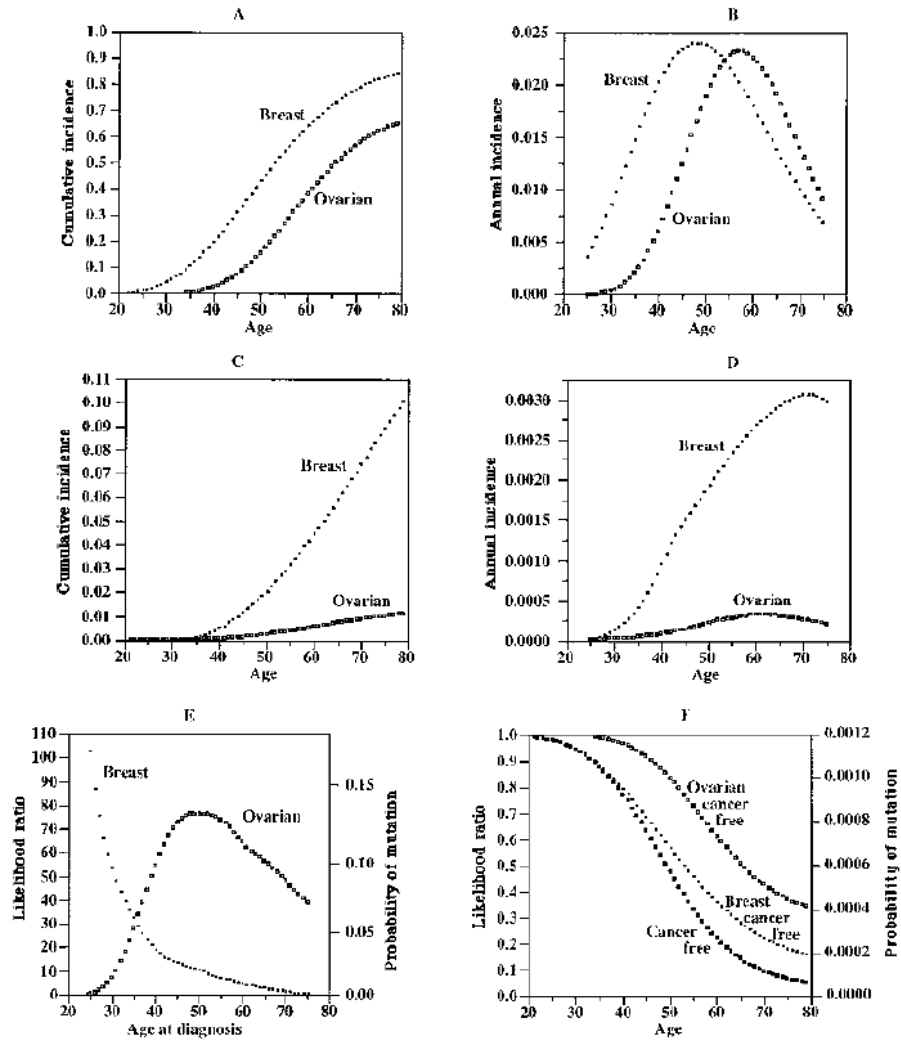


Figura 2.3: Riesgo acumulado para portadores (A) y no portadores de mutación (C), incidencia específica por edad para ambos grupos (B y D), y likelihood ratio y probabilidad a posteriori de portar la mutación en paciente con enfermedad desarrollada (E) y paciente sin enfermedad desarrollada (F). Tanto para CM como para CO [7].

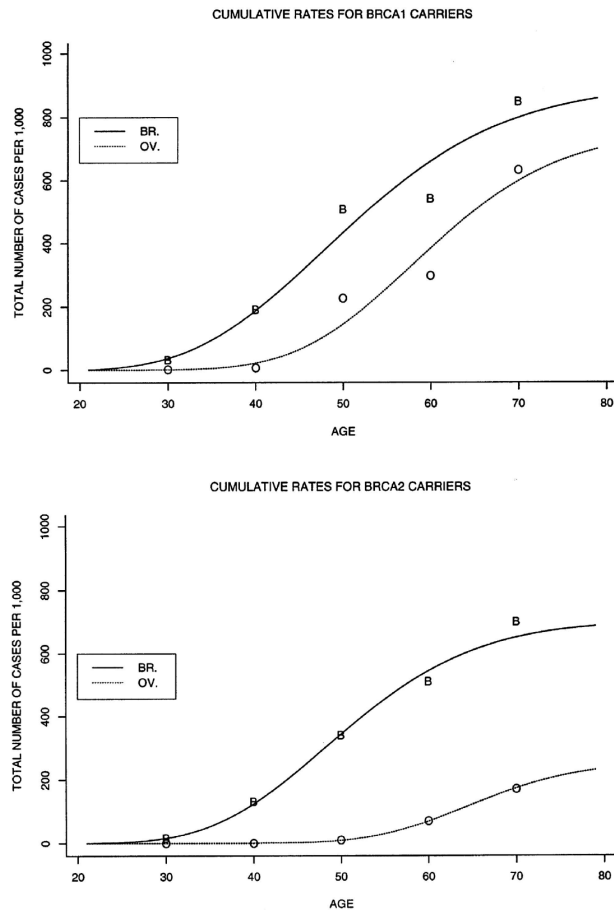


Figura 2.4: Tasas acumuladas para CM y CO en portadores de BRCA I (arriba) y BRCA II (abajo) [8]

2.4. Modelo Tyrer-Cuzick o IBIS

El modelo Tyrer-Cuzick incluye el conjunto más completo de variables para el cálculo de riesgo. Es el único que integra información genética, historia familiar y factores personales y reproductivos, permitiendo estimar de manera global el riesgo tanto para poblaciones con el gen BRCA mutados como para las que no poseen la mutación [9]. Además, a partir de la hipótesis de que los genes BRCA I y II no explican en su totalidad la prevalencia de la enfermedad entre los miembros de una familia, el modelo incluye además genes de baja penetrancia para el cálculo de riesgo.

Se hizo uso del modelo dominante para la estimación de riesgo genético, donde se tomó un modelo de dos locus, uno para el BRCA I y BRCA II y otro para un gen hipotético de menor penetrancia, que representa varios posibles genes de baja penetrancia en el CM. Para simplificación de cálculos, se asumió que:

- los genes BRCA I y II se hallan en un mismo locus,
- si una persona posee mutación en el gen BRCA I y II se lo considera con mutación en gen BRCA I únicamente y
- el gen hipotético es de carácter dominante, por lo tanto, ser homocigota o heterocigota conlleva al mismo fenotipo

Como consecuencia, el modelo contempla seis fenotipos que son la combinación entre el locus normal, mutado en BRCA I o mutado en BRCA II, y el locus del gen hipotético normal o mutado. Cada uno de estos seis fenotipos posee su propia distribución de probabilidad de desarrollar CM [9].

La expresión que utilizaron para el cálculo de probabilidad de desarrollar CM fue la siguiente :

$$P(cancer) = 1 - \left(1 - \sum_{i=1}^6 p_i F_i(t_1, t_2) \right)^\alpha$$

En la ecuación, p_i es la probabilidad de poseer el fenotipo i de los seis posibles. Esta probabilidad se calcula de forma análoga al modelo de BRCA-PRO, utilizando el teorema de Bayes [42].

$F_i(t_1, t_2)$ es la probabilidad de contraer CM durante las edades t_1 y t_2 dado el fenotipo i y esta se calcula a partir de la diferencia entre las funciones de supervivencia basal respecto al fenotipo correspondiente. Estas funciones de supervivencia se obtuvieron a partir de métodos numéricos (Newton-Raphson), y considerando tres funciones S_i para cada estado del BRCA, y

para cada franja de edades y para los posibles genotipos en el gen hipotético. En las Tablas 2.9 y 2.10 se observan las expresiones para cálculo de riesgo de los seis fenotipos posibles del modelo y algunos valores de riesgo acumulado, respectivamente.

Fenotipos	Descripción	Probabilidad de desarrollar CM entre edades t_i a t_j
1	Sin BRCA, sin gen hipotético	$S_0(t_i) - S_0(t_j)$
2	Sin BRCA, con gen hipotético	$S_0(t_i)^\theta - S_0(t_j)^\theta$
3	Con BRCA I, sin gen hipotético	$S_1(t_i) - S_1(t_j)$
4	Con BRCA I, con gen hipotético	$S_1(t_i)^\theta - S_1(t_j)^\theta$
5	Con BRCA II, sin gen hipotético	$S_2(t_i) - S_2(t_j)$
6	Con BRCA II, con gen hipotético	$S_2(t_i)^\theta - S_2(t_j)^\theta$

Tabla 2.9: Los seis fenotipos con sus respectivas expresiones de riesgo [9]

Edades	Riesgo acumulado para BRCA I ($1 - S_1(t)$)	Riesgo acumulado para BRCA II ($1 - S_2(t)$)
30	0.036	0.06
40	0.18	0.12
50	0.49	0.28
60	0.64	0.48
70	0.71	0.84

Tabla 2.10: Estimaciones de riesgo acumulado para portadores de mutaciones [9]

El último parámetro de la ecuación es α , y representa el riesgo relativo atribuido a los factores de riesgo no hereditarios. Éstos son los enumerados a continuación:

- Tardía edad de menarca
- Nula cantidad de partos
- Tardía edad de primer nacimiento
- Tardía edad de menopausia
- Diagnósticos de hiperplasias atípicas
- Diagnósticos de carcinoma lobulillar in situ

- Altura
- Alto índice de masa corporal

Se asumió un modelo proporcional de riesgo, donde estos RR se multiplican. La edad de menarca decrece el riesgo en un factor de 0,95 por cada año que avanza, mientras que la edad de menopausia lo aumenta en 1.028. La presencia de hiperplasias atípicas aumenta en un factor de 4, y la de un carcinoma lobulillar in situ en 8. El resto de los factores pueden observarse en la Tabla 2.11

Variable	Riesgo relativo
Edad de primer nacimiento	
Sin partos	1.0
< 20	0.67
20 - 24	0.74
25 - 29	0.88
> 30	1.04
Altura (m)	
< 1.6	1.0
1.6 - 1.7	$1.05 + 2 \times (\text{altura} - 1.6)$
> 1.7	1.24
IMC (kg/m^2)	
< 21	1.0
21 - 23	1.14
23 - 25	1.15
25 - 27	1.26
> 27	1.32

Tabla 2.11: Riesgos relativos del modelo de IBIS [9]

Finalmente, ya habiendo determinado las variables del modelo, se presenta en la Figura 2.5 el flujograma para proceder al cálculo de riesgo en pacientes.

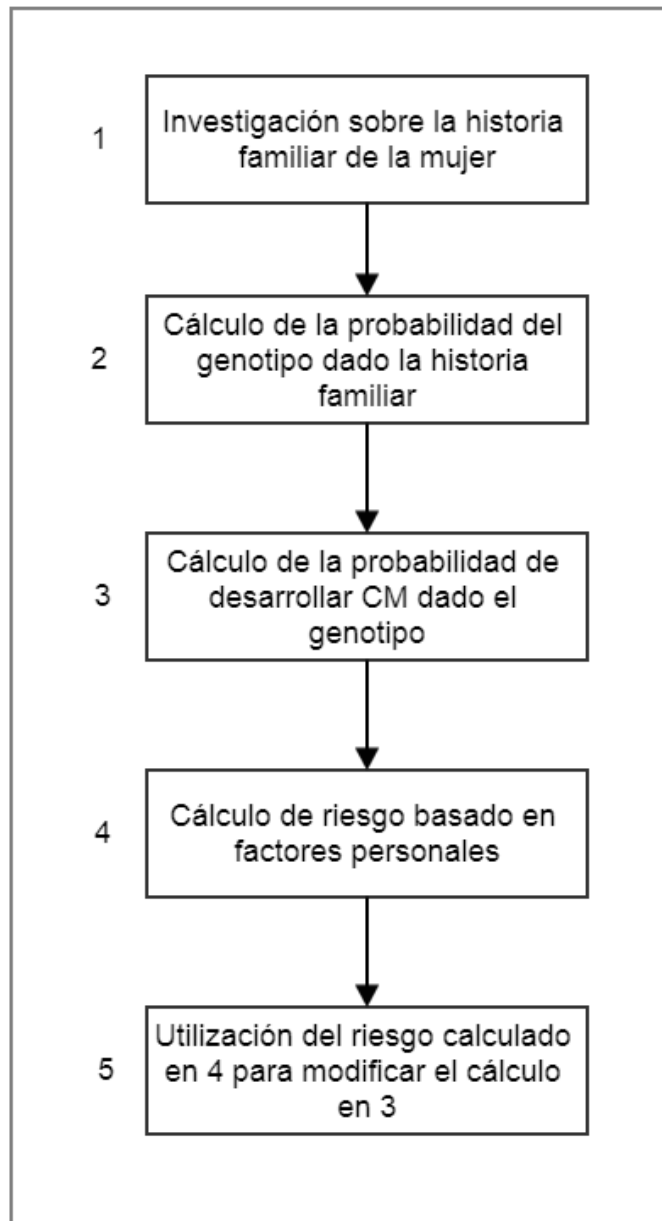


Figura 2.5: Proceso de cálculo de riesgo en pacientes según el modelo IBIS [9].

Capítulo 3

Selección de modelo de evaluación

En el Capítulo 2 se llevó a cabo la descripción de cuatro modelos para evaluación de riesgo de CM, el Gail, Cluas, BRCAPRO y Tyrer-Cuzick. A continuación, se evalúan y comparan las diferentes características de cada uno para determinar el modelo a desarrollar en el CDSS. Se analizan los desempeños de cada uno, las validaciones en diferentes poblaciones, las variables que contemplan y los contextos de mayor eficacia.

Existen dos parámetros que se tomaron en cuenta para cuantificar la eficacia de los modelos. Uno es la calibración que refleja el desempeño global, y se lo calcula como la cantidad de enfermos estimados por el modelo sobre la cantidad de enfermos observados en la población analizada (E/O). Si E/O es menor a 1, el modelo subestima el riesgo. Caso contrario, hay una sobrestimación. El segundo parámetro es la discriminación, es decir, la capacidad de predecir el riesgo a nivel individual. Para ello se calcula el área bajo la curva (AUC) característica operativa de receptor (ROC), que representa gráficamente la sensibilidad contra la especificidad. Si $AUC = 1$, la prueba tiene una capacidad óptima de discriminación. Si $AUC = 0.5$, la discriminación del modelo es nula.

Como primer caso, en el estudio de Amir [10], se calcularon estos dos parámetros para los cuatro modelos estudiados. En las Tablas 3.1 y 3.2 se pueden observar los resultados. También se muestra la curva ROC en la Figura 3.1.

Modelo de evaluación de riesgo	Área	Intervalo de confianza (95 %)	
		Límite inferior	Límite superior
Gail	0.735	0.666	0.803
Claus	0.716	0.648	0.784
BRCAPRO (Ford)	0.737	0.671	0.803
Tyrer-Cusick	0.762	0.700	0.824

Tabla 3.1: Áreas bajo la curva ROC para modelos de Gail, Claus, BRCAPRO y Tyrer-Cuzick según estudio de Amir [10].

Modelo de evaluación de riesgo	Observado	Esperado	E/O	IC 95 %
Gail	64	44.3037	0.69	0.54 to 0.90
Claus	64	48.5565	0.76	0.59 to 0.99
BRCAPRO (Ford)	64	42.2790	0.66	0.52 to 0.86
Tyrer-Cuzick	64	69.5653	1.09	0.85 to 1.41

Tabla 3.2: Cantidad de casos esperados vs. observados para los 4 modelos ($n = 3150$) [10]

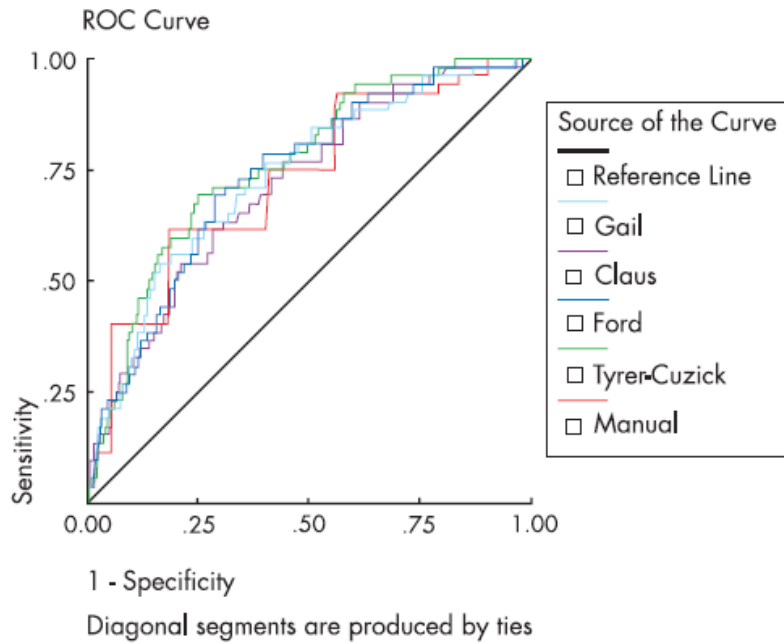


Figura 3.1: Curvas ROC para los 4 modelos [10]

Si bien varios estudios han validado el modelo de Gail en poblaciones

independientes [13, 30, 32, 43], se puede concluir del estudio de Amir que el modelo de Tyrer-Cuzick es el de mejor desempeño tanto a escala global como individual, y que los resultados de los otros tres modelos subestiman el riesgo absoluto.

Sin embargo, existen dos problemas para la implementación del modelo IBIS. Por un lado la disponibilidad de la información necesaria para el cálculo de riesgo. En la Tabla 3.3, que muestra las variables contempladas en cada modelo, se puede observar que de los cuatro, éste es el que mayor volumen de datos requiere para realizar el cálculo eficaz de riesgo. Si bien se pueden obtener resultados del modelo aún con ausencia de datos, la precisión disminuye. Además, en caso de que la información no estuviera disponible en la historia clínica del paciente, y se decidiera registrarla e ingresarla durante una consulta clínica, el tiempo que demoraría sería demasiado en el marco de la consulta y por ende sería poco practicable. Por otro lado, si bien algunos estudios, como el de Amir, sugieren que el modelo de IBIS es el de mayor eficacia, todavía es necesario una validación más rigurosa en diferentes poblaciones para asegurar su buen desempeño.

Caso contrario es el modelo de Gail, que si bien en varios estudios demuestran una menor eficacia con respecto al modelo de Tyrer-Cuzick, éste ha sido validado en numerosas poblaciones (Figuras 3.2 y 3.3) y por ende, como se mencionó en la sección anterior, es el más difundido y el más estudiado. Ante la posible implementación en una población cosmopolita como la de la Ciudad Autónoma de Buenos Aires, esta característica adquiere importancia. Además, el volumen de información que requiere para el cálculo es significativamente menor: 6 variables para Gail contra 14 variables para IBIS (Tabla 3.3).

Los modelos de Claus y BRCAPRO no poseen tantas validaciones en poblaciones independientes, y además están limitados a un subconjunto de pacientes que son los que conllevan historia familiar con CM y/o CO. Estos modelos, si bien pueden ser eficaces para calcular el riesgo en estos casos, no cumplen con el objetivo de identificación de pacientes de alto riesgo en la población en general, establecido en el Capítulo 1. Éstos modelos dejan de lado todos los factores de riesgo no hereditarios (hormonales, reproductivos o clínicos) que pueden desencadenar la enfermedad, y por ende, por lo específico, no contemplan otros subconjuntos que son relevantes para un análisis poblacional.

Como contraparte, el modelo de Gail considera, aunque parcialmente, la historia familiar. Al incorporar únicamente los familiares de primer grado entre sus variables, no incluye el linaje paternal, posible origen de mutación en alelo del gen BRCA. Tampoco contempla antecedentes de CO, de significativa correlación con el desarrollo de CM [44]. Para estos casos el modelo puede ser

ineficaz, y es por eso que en el árbol de decisión de la Figura 3.4, el modelo de Gail se utiliza cuando:

- Posee en la historia familiar sólo afectados de primer grado por CM y
- Sin casos en la familia, posee factores de riesgo hormonales y/o reproductivos, pero no antecedentes patológicos benignos [45]

Como conclusión, entendiendo que no hay un modelo que pueda abarcar todos los casos de manera óptima, y considerando los estudios de validación (aunque aún no validado en poblaciones hispanas), el volumen de información requerido y las ventajas de aplicación en la población en general, se eligió el modelo de Gail para desarrollar el CDSS correspondiente a la segunda parte de este trabajo. Su implementación permitirá al profesional tomar la decisión de llevar a cabo un monitoreo intensivo, iniciar tratamientos quimiopreventivos, realizar posibles mastectomías, o simplemente aconsejar respecto de ciertos cuidados a la paciente.

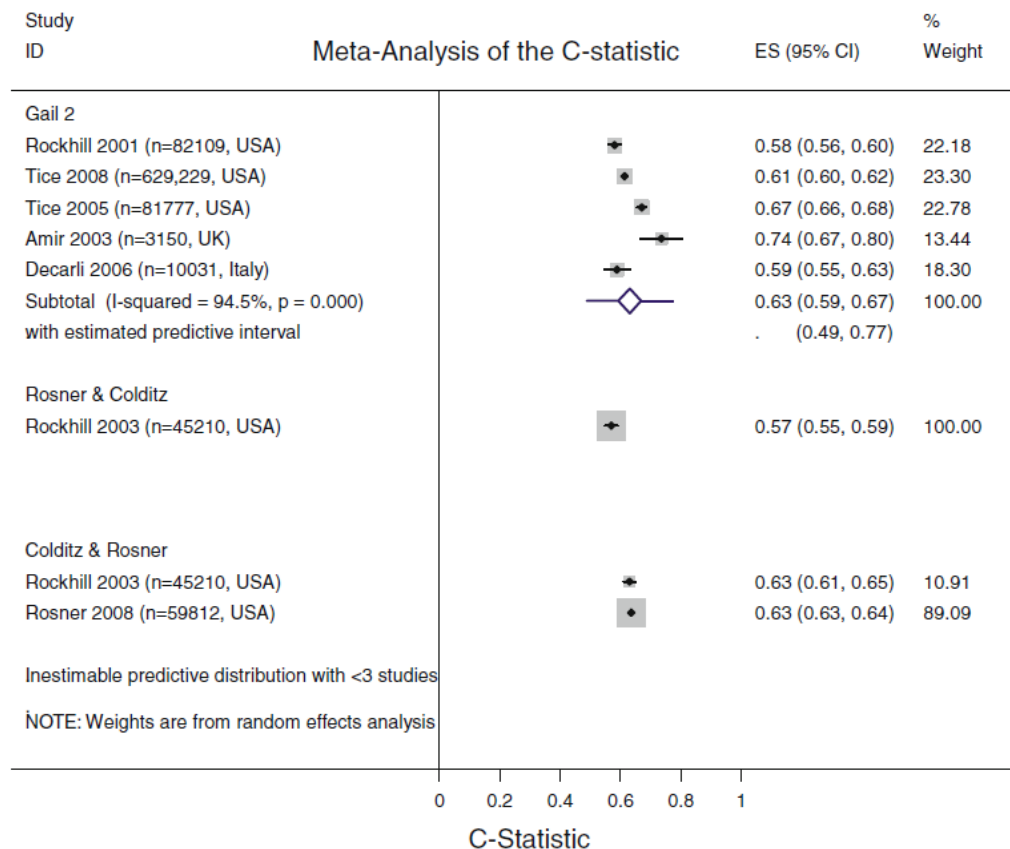


Figura 3.2: Valores de AUC para distintas validaciones del modelo de Gail [11]

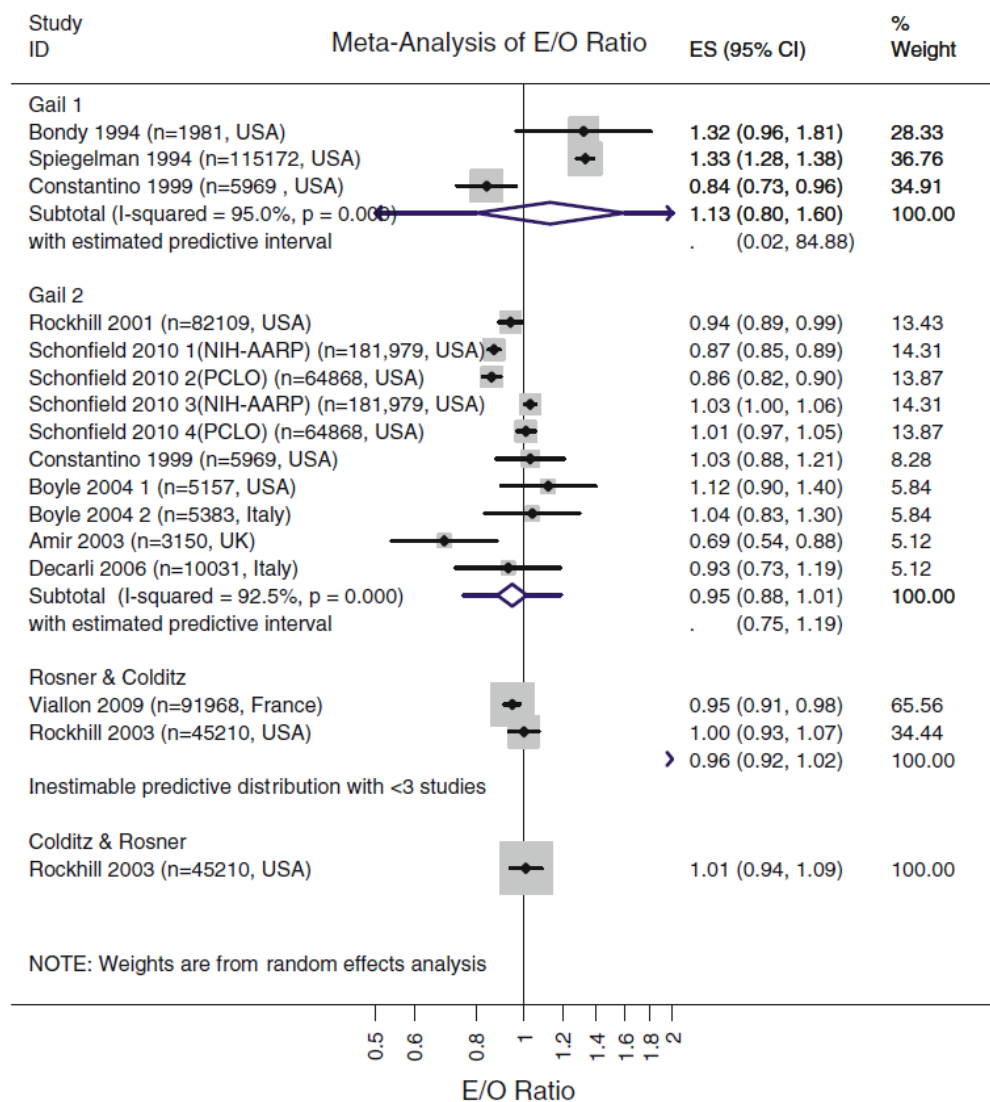


Figura 3.3: Valores de E/O para distintas validaciones del modelo de Gail [11]

Factor de riesgo	Gail	Claus	BRCAPRO	Tyrer-Cuzick
Información personal				
Edad	Sí	Sí	Sí	Sí
Índice de masa corporal	No	No	No	Sí
Consumo de alcohol	No	No	No	No
Factores hormonales/re-productivos				
Edad de menarca	Sí	No	No	Sí
Edad de primer parto	Sí	No	No	Sí
Edad de menopausia	No	No	No	Sí
Uso de terapia de reemplazo hormonal	No	No	No	Sí
Uso de pastillas anticonceptivas	No	No	No	No
Lactancia materna	No	No	No	No
Estrógeno plasmático	No	No	No	No
Patologías de mama				
Biopsias de mama	Sí	No	No	Sí
Hiperplasia atípica ductal	Sí	No	No	Sí
Carcinoma in situ lobulillar	No	No	No	Sí
Densidad de tejido mamario	No	No	No	No
Historia familiar				
Familiares de primer grado	Sí	Sí	Sí	Sí
Familiares de segundo grado	No	Sí	Sí	Sí
Familiares de tercer grado	No	No	No	No
Edad de comienzo de CM	No	Sí	Sí	Sí
CM bilateral	No	No	Sí	Sí
CO	No	No	Sí	Sí
CM en hombres	No	No	Sí	No

Tabla 3.3: Variables consideradas en distintos modelos de evaluación de riesgo [16]

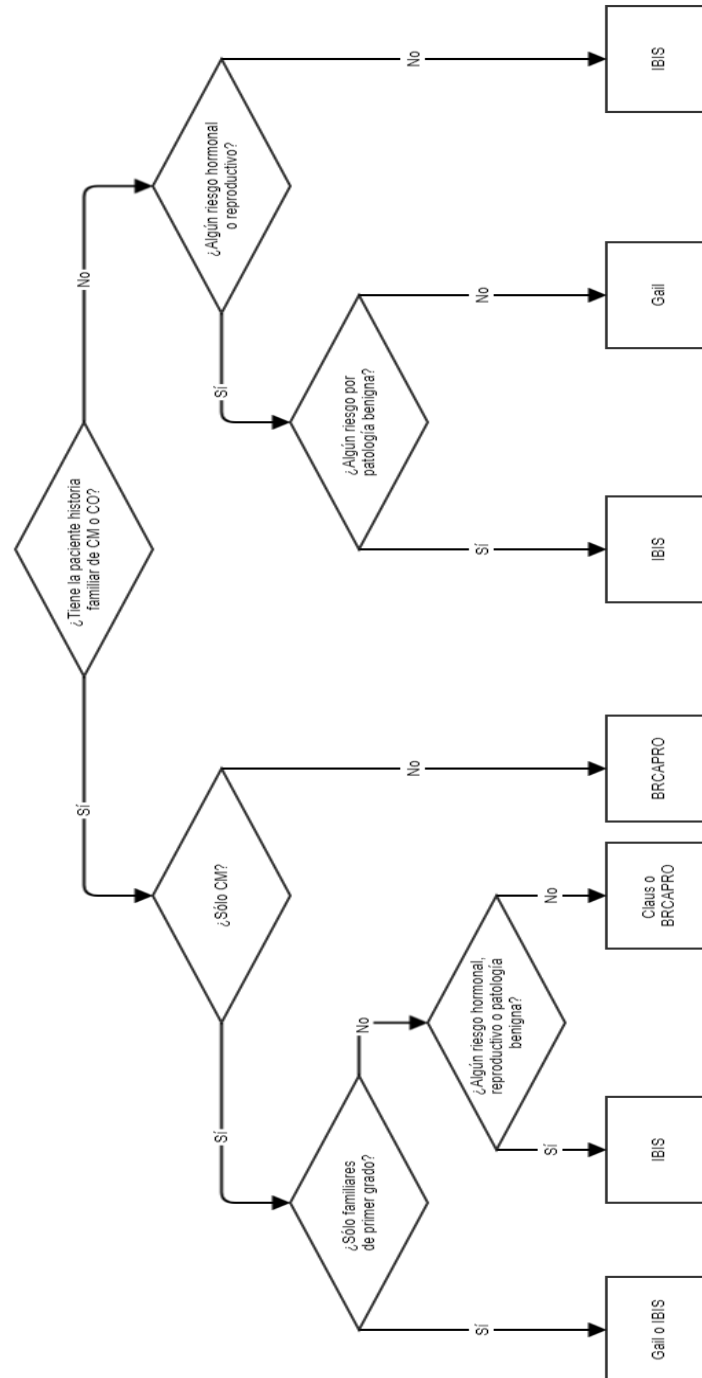


Figura 3.4: Árbol de decisión para elección de modelo de evaluación de riesgo de CM. [4]

Parte II

Diseño y desarrollo de CDSS basado en modelo de Gail

Capítulo 4

Etapas de vida del CDSS

En esta segunda parte del proyecto se plasma el proceso de desarrollo del CDSS. Para ello se adoptó la metodología en cascada de las 5 etapas del ciclo de vida (Figura 4.1), obteniendo como producto final el programa para cumplir con el objetivo establecido en el inicio.

Estas 5 etapas de desarrollo constan de un primer análisis de requerimientos en donde se evalúan las necesidades y consecuentemente se establecen las directrices que regirán todo el ciclo de vida. La etapa posterior es el diseño donde se definen la arquitectura, los módulos, el flujo de la información. También se plasma la estructura del código y de los algoritmos que se desarrollarán. Luego, la fase de implementación es la propia generación de código fuente, ya definido previamente el lenguaje, el entorno de programación, las librerías a utilizar. Posteriormente, la fase de prueba en donde se ensambla el sistema, se realiza la puesta en marcha y las primeras interacciones sistema - usuario final. Por último, el mantenimiento es la etapa donde las no conformidades del usuario y las nuevas sugerencias que puedan surgir de la experiencia son implementadas en el programa. Es por eso la importancia del lazo de retroalimentación en esta metodología de desarrollo en cascada, para lograr la mejora continua.

En este capítulo se describirán estas etapas a lo largo del ciclo de vida del CDSS.

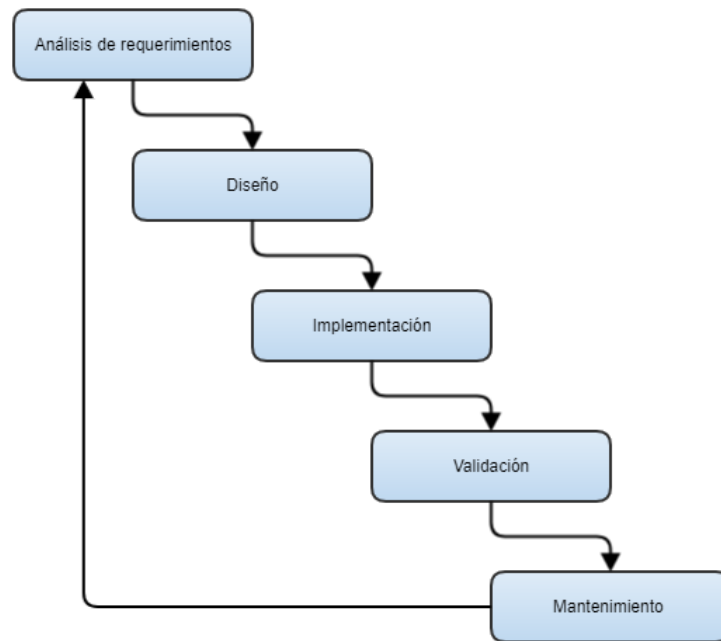


Figura 4.1: Ciclo de vida de software. Modelo de cascada realimentado.

4.1. Requerimientos

Informar cuantitativamente a profesionales de la atención primaria a la salud acerca del riesgo que posee una paciente de desarrollar CM, conociendo a priori los factores de riesgo, es el requerimiento principal del CDSS. El objetivo es la identificación de pacientes con alto riesgo tal como se planteó en el Capítulo 1.

Durante el análisis de los modelos, se observó que existen varios subconjuntos poblacionales dentro de las potenciales pacientes según sus factores de riesgo, y que las eficacias de los modelos predictivos varían en función de estos grupos. Por las razones mencionadas en el Capítulo 3, el modelo de Gail es el que se utilizará para la evaluación cuantitativa de riesgo.

Los factores de riesgo que deberá tomar en cuenta el CDSS son los analizados en la sección 2.1, AGE CAT, NUMREL, AGEFLB, AGEMEN y NBIOPS, que se observan en la ecuación 2.1. A estas 5 variables se le adiciona la detección de hiperplasias atípicas de mama (AH) que no se encuentra incluido en el modelo original de Gail. Los RR asociados a este factor de riesgo se encuentran en la Tabla 4.1.

En esta primera instancia de desarrollo del programa, la información deberá ser ingresada por el usuario. Los usuarios finales serán principalmente profesionales médicos de atención primaria, para que puedan ser asistidos en la toma de decisión para una eventual interconsulta con el especialista idóneo.

Factor de riesgo AH	RR
Sin biopsias	1.00
Con biopsias, sin detección de AH	0.93
Con biopsias e AH detectada	1.82

Tabla 4.1: Riesgos relativos asociado a detección de hiperplasias atípicas (AH) [17]

4.2. Arquitectura y diseño

En función del modelo de Gail, la arquitectura del CDSS de la Figura 1.1 toma la configuración de la Figura 4.2.

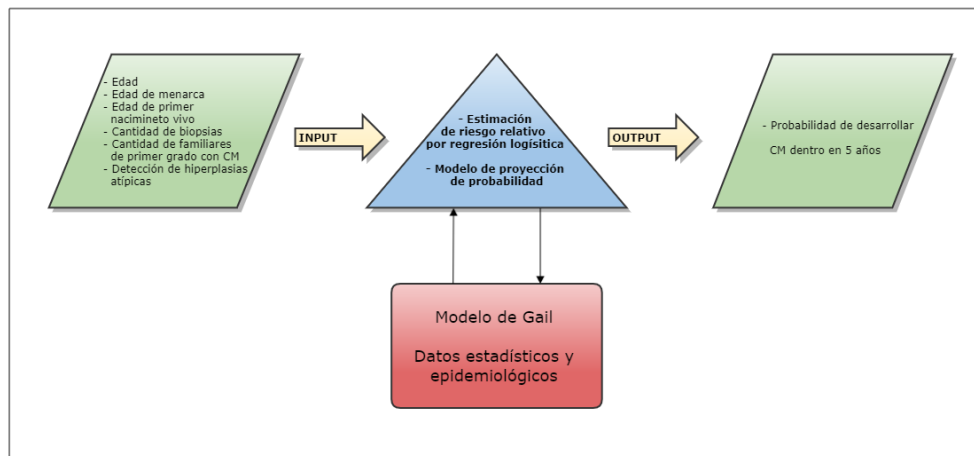


Figura 4.2: Arquitectura de CDSS de evaluación de riesgo de CM

Las 6 variables de entrada son las mencionadas anteriormente. La estimación por regresión logística del RR se resume al producto entre los RR asociados a cada uno de los factores de riesgo (ecuación 2.2). Una vez calculado el riesgo del paciente, se procede a calcular la proyección a 5 años

a partir de la ecuación 2.4. Como resultado, el CDSS devuelve la probabilidad de desarrollar CM en 5 años asumiendo que los factores de riesgo en la persona no cambiarán durante ese período de tiempo. Tal como se menciona en la Figura 1.4, si dicha probabilidad supera el 1,67 %, según el modelo se trata de una paciente de alto riesgo. Si esta probabilidad cumple con la condición, entonces se genera una advertencia para evaluar interconsulta con el especialista, quien será el que realice el seguimiento, prescriba los estudios complementarios necesarios y determine los tratamientos o medidas preventivas que se llevarán a cabo.

Para implementar el CDSS se optó por el desarrollo de un servicio web, de manera tal que pueda ser utilizado por múltiples clientes simultáneamente. Paralelamente, para la interfaz con el cliente y para la carga manual de entradas, se propuso una aplicación web.

4.2.1. Servicio web

Según la World Wide Web Consortium los servicios web son aplicaciones con capacidad de interoperar con la web y que intercambian datos con el fin de ofrecer servicios a los usuarios a través de la web [46].

Mediante protocolo SOAP, estos servicios intercambian información estructurada en XML, lo que permite la interoperabilidad y escalabilidad.

La arquitectura estándar para el servicio web del CDSS se la puede observar en la Figura 4.3, donde los clientes pueden realizar las consultas a través de la aplicación web, un servidor web aloja el programa de cálculo de riesgo y otro servidor funciona como base de datos. Estas dos instancias de alojamiento, tanto del procesamiento como de la información, no se contemplan en este proyecto, quedando el programa almacenado en el servidor local.

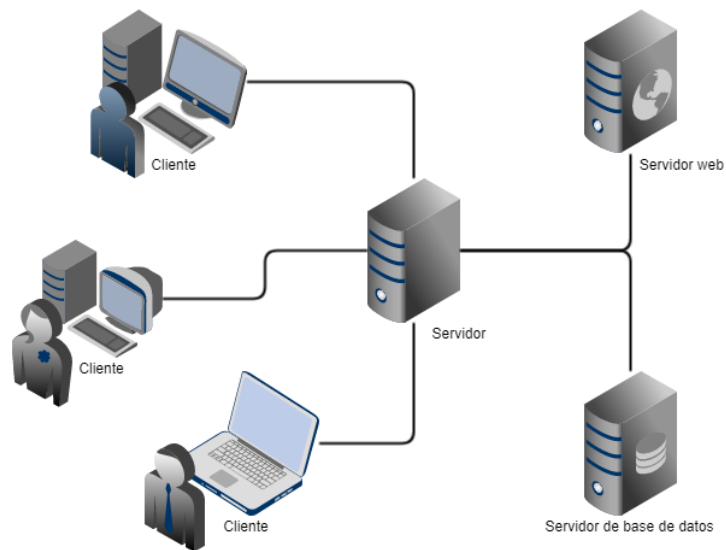


Figura 4.3: Arquitectura de servicio web

4.2.2. Diseño de programa

Para el desarrollo del servicio web se utilizó la plataforma ASP.NET de Microsoft, mediante el IDE Visual Studio, programado en lenguaje C#. Este entorno ofrece soporte con librerías que permiten la abstracción respecto a la codificación de protocolos, lo que facilita significativamente la etapa de implementación.

En particular se utilizaron las clases `WebService` para el desarrollo del servicio web, donde se definen las modalidades de intercambio de información y los métodos, y la clase `Page` para la aplicación, donde se incluyen los eventos, representadas en archivos `.asmx` y `.aspx` respectivamente. También se generó la clase `Comunicacion` para gestionar las variables de entrada y salida.

El flujograma de la información en el proceso de soporte a la decisión y el diagrama de módulos del sistema están representados en las Figuras 4.4 y 4.5.

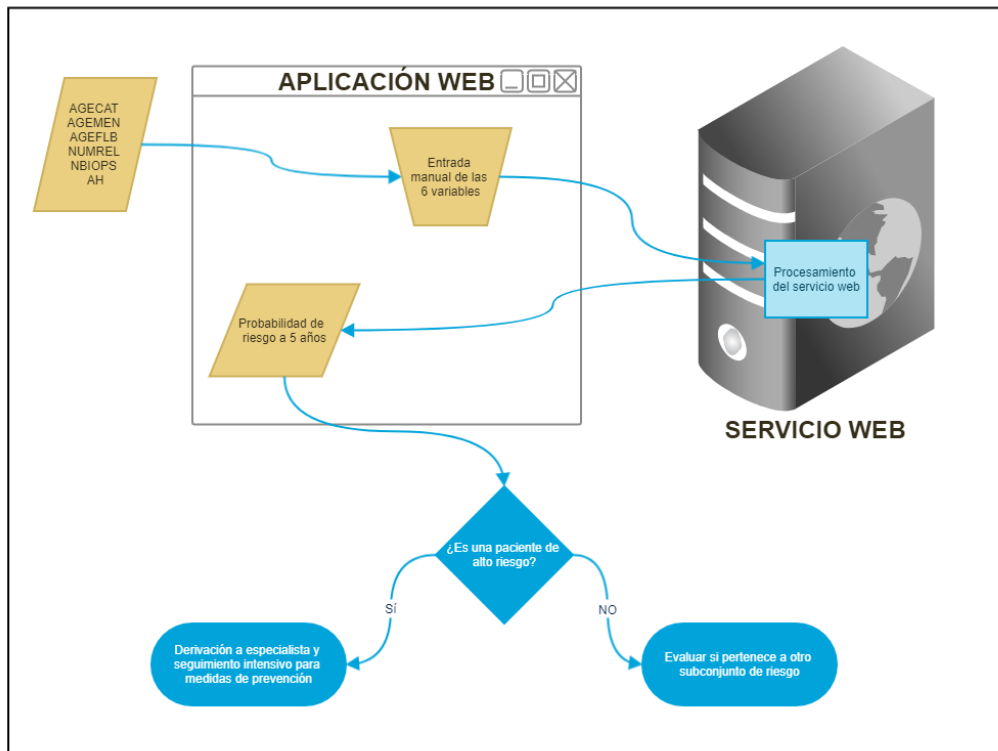


Figura 4.4: Flujograma del proceso de soporte a la decisión

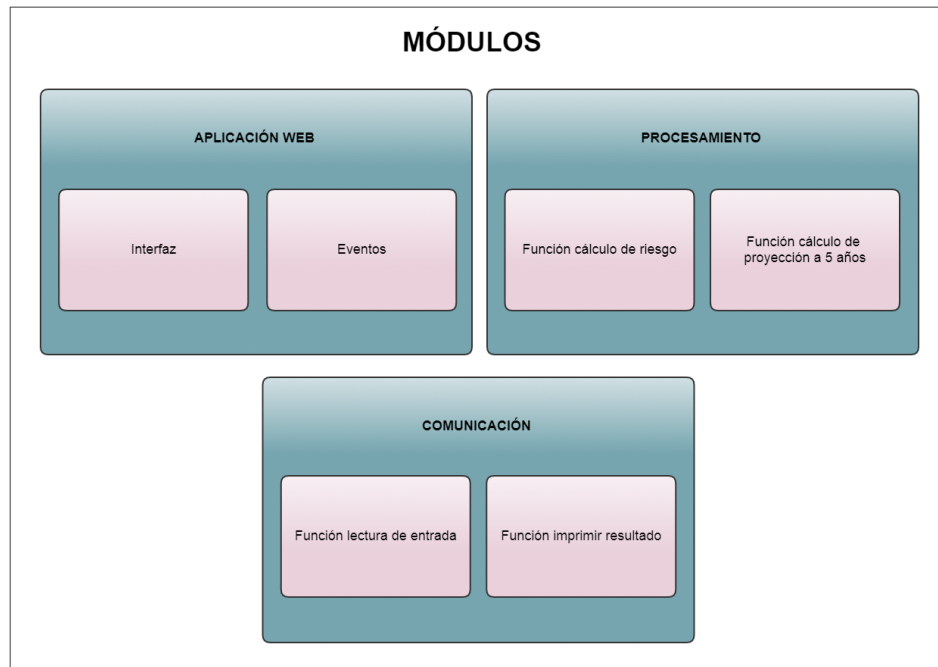


Figura 4.5: Módulos del CDSS

4.2.3. Interoperabilidad y estándares HL7

La interoperabilidad es la capacidad de dos o más sistemas de intercambiar información –interoperabilidad sintáctica - y hacer uso de ella – interoperabilidad semántica [47].

En el ámbito de la informática médica, la definición de interoperabilidad según la Healthcare Information and Management Systems Society (HIMSS) se remite a “...la capacidad de diferentes sistemas de información en salud (sistemas hospitalarios, departamentales, registros clínicos electrónicos, etc.) para intercambiar datos y usar la información que ha sido intercambiada dentro y a través de los límites de la organización, con el fin de mejorar la prestación efectiva de los cuidados de salud a individuos y comunidades...” [48]. Los beneficios de la interoperabilidad pueden analizarse desde varias aristas. La calidad de la prestación de los servicios de salud es una de ellas. En este sentido es fundamental el acceso a información completa, precisa y actualizada por parte del profesional médico para tomar las decisiones clínicas que correspondan en el lugar y en el momento que sea necesario, y a la vez

evitar la incidencia de errores médicos. El hecho de que la información clínica esté segmentada en silos, duplicada, incompleta, no disponible, son factores que indefectiblemente impactan en la calidad de la salud.

Los estándares Health Level Seven o HL7 fueron creados para permitir la interoperabilidad entre los sistemas de información clínica, es decir el intercambio y la integración de la información [49]. En particular, FHIR es un conjunto de recursos de estructuración de la información clínica tanto en XML como en JSON, creado por la organización HL7. Estos recursos son un proyecto de estándar que facilita notablemente la implementación de interoperabilidad sintáctica y semántica.

Si bien escapa a los objetivos del proyecto la aplicación de los protocolos HL7 en el desarrollo del CDSS, es relevante subrayar la importancia de que la aplicación pueda ser interoperable. Aún así, se deja planteado, como ejemplo, el caso de las variables de salida del CDSS, donde existe un recurso FHIR para riskassessment aplicable (Apéndice A). En particular, en el código JSON el campo RiskAssessment.prediction.probability, contendría el valor de la probabilidad obtenida, en RiskAssessment.prediction.when la proyección a 5 años, en RiskAssessment.method se mencionaría el modelo de Gail, y así con toda la información del proceso de evaluación de riesgo.

De la misma forma, para cada variable de entrada existen recursos y campos asociados que pueden ser analizados fácilmente para lograr interoperabilidad.

4.3. Implementación

4.3.1. Módulo comunicación

En este módulo, la función de lectura de entradas, readVariables() según código del Apéndice B, tiene como entradas los campos completados por el usuario en la interfaz, las seis variables de entrada. Para evitar errores en esta etapa, la ingreso de datos está estructurado por la clase dropdownlist, limitándolo a las variables que son aceptadas por el modelo. Así, el usuario no podrá ingresar datos que no sean interpretables en el posterior procesamiento, reduciendo así el riesgo de errores durante la ejecución.

La segunda función setOutput() establece la probabilidad a 5 años, y en caso de que dicho resultado pronostique un alto riesgo en la paciente, también se genera el mensaje de advertencia (Figura 4.6).

El desarrollo de un módulo aparte para el manejo de las entradas y salidas es estratégico para una eventual aplicación de protocolos HL7. De esta manera, no es necesario modificar el código de los otros módulos.

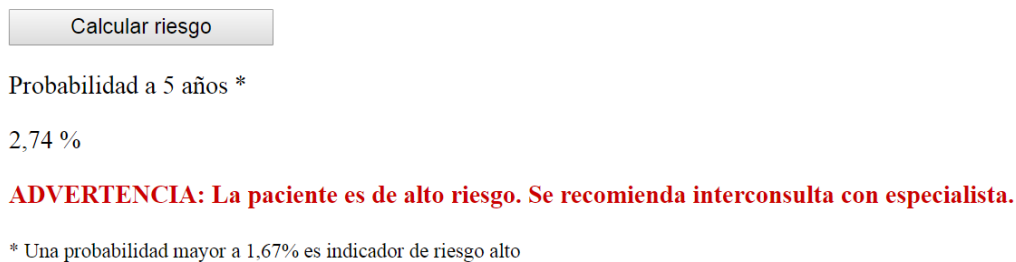


Figura 4.6: Mensaje de advertencia

4.3.2. Interfaz de usuario

La aplicación web establece la interfaz de usuario (Figura 4.7) y a la vez es donde se invocan las funciones del servicio web y del módulo de comunicación cuando el usuario decide realizar el cálculo con un click el botón Calcular riesgo. Obsérvese que el campo de hiperplasias atípicas no se encuentra habilitada. Esto se debe a que la única forma realizar este diagnóstico es a partir de una biopsia. Si la paciente no tiene biopsias hechas, es decir, que el campo Cantidad de biopsias se encuentra en 0, no es posible que se le haya diagnosticado hiperplasia atípica, y por ende, el se inhabilita y se la presupone con valor negativo.

En cuanto a la implementación del evento, de breve sintaxis, consta de cuatro partes. La primera donde se declaran los dos objetos a utilizar, el WebService y el Communication. La segunda es la inicialización de los atributos de entrada del objeto Communication en función de los datos ingresados por el usuario en los dropDownList respectivos. La tercera es el procesamiento de estas variables, donde interviene la clase WebService con sus dos métodos. Y la cuarta define el estado de los atributos de salida del objeto Communication.

← → localhost:63152/WebForm1.aspx ☆ P ↗ ⋮

Aplicaciones

MODELO DE GAIL

Cálculo de riesgo de cáncer de mama

Edad

Edad menarca

Edad de primer parto

Cantidad de parientes de primer grado con cáncer de mama

Cantidad de biopsias de mama

Detección de hiperplasias atípicas

Probabilidad a 5 años *

* Una probabilidad mayor a 1,67% es indicador de riesgo alto

Figura 4.7: Interfaz de usuario

4.3.3. Módulo procesamiento

La función de cálculo de riesgo devuelve el riesgo relativo de la paciente, determinado por 4 factores de riesgo relativo en función de las 6 variables de entrada. Éstas se las determina a partir de sentencias condicionales anidadas (Apéndice C), según las Tablas 2.1 y 4.1.

La segunda función de proyección de probabilidad, plasma la ecuación 2.4. Toma como variables de entrada la edad de la paciente, y los valores de riesgo de la persona en la actualidad y dentro de 5 años suponiendo que los factores de riesgo no cambian. Estas dos últimas son la salida de la función anterior.

Para la implementación se tomaron las incidencias de CM por edad de la SEER, $h_1(t)$ y los riesgos de muerte por otras causas que no sean CM, $h_2(t)$. La distribución de los valores de ambos parámetros es constante para cada franja etaria (Tabla 2.2), 13 intervalos de 5 años, en el rango de 35 a 90. Debido a que no todos los casos de desarrollo de CM suceden en la población expuesta a los factores de riesgo considerados por el modelo, el primer parámetro se lo ajusta con un factor. Este factor de ajuste es constante para todas las edades y es 1 menos el riesgo atribuibles ¹. Se considerará 0.5788 para este valor [13].

En cuanto al cálculo de la probabilidad, la sumatoria de la ecuación 2.4

¹En epidemiología, el riesgo atribuible refleja la diferencia en una condición entre la población expuesta y la no expuesta.

se reduce a dos términos cuando la proyección es a 5 años. El primero corresponde a la franja etaria de la edad de la persona, y el segundo término a la franja siguiente. Los numeradores de los términos de supervivencia sólo aparecen en este segundo término. Estos dos términos están expresados en el Apéndice D.

4.4. Verificación del algoritmo

La verificación es el proceso mediante el cual se justifica que el algoritmo cumple con su requerimiento funcional. Para llevar a cabo la prueba del programa de cálculo de riesgo basado en el modelo de Gail se exploró aleatoriamente el espacio de estado posibles y las salidas fueron comparadas con los resultados obtenidos por MDCalc [50], servicio de soporte a la decisión clínica disponible en la web.

Los resultados obtenidos por el programa y por el servicio MDCalc fueron coincidentes en todos los casos, permitiendo inferir la efectividad del algoritmo evaluado.

Para el desarrollo del algoritmo, se consideró una población de raza blanca, como en el modelo original. Esta variable está incluida en el servicio de MDCalc, y pueden considerarse diferentes razas como la asiática, africana, hispana. La razón por la cual se fijó esta variable en el programa se debe a que el modelo aún no se encuentra validado en la población argentina, y por ende, la discriminación entre razas no otorga valor descriptivo real.

4.5. Validación y puesta en marcha

El modelo de Gail no ha sido validado aún en la población Argentina, tal como se mencionó en la sección 2.1.1. Será el próximo desafío que se llevará a cabo, por fuera del alcance de este proyecto, junto al Departamento de Informática en Salud del Hospital Italiano. A su vez especialistas oncológicos y en diagnóstico por imágenes participarán y acompañarán este proceso.

Los datos clínicos para la validación podrán ser simulados, o, en caso de ser datos reales contarán con aprobación del Comité de Ética y Protocolos de Investigación del Hospital Italiano de Buenos Aires.

Los indicadores que se tomarán en cuenta serán los mismo mencionados en el Capítulo 3. Por un lado la curva ROC para evaluar la capacidad de discriminación, y por otro la tasa E/O para observar la calibración del modelo. Una vez superada esta etapa se dará inicio a las primeras pruebas piloto con profesionales médicos para analizar la experiencia de usuario, la integración

con la historia clínica electrónica del hospital y consecuentemente realizar las mejoras que sean menester.

Capítulo 5

Conclusión y desafíos a futuro

La motivación de este proyecto final es la alarmante incidencia y mortalidad del cáncer a nivel nacional y mundial. Es la segunda causa de muerte en el planeta. En particular, el CM en nuestro país, es la mayor causa de muerte en mujeres, entre las muertes provocadas por cáncer. Es por eso que la directriz de este proyecto es la implementación de un sistema para evaluación de riesgo que asista a la toma de decisión clínica y de esa manera contribuya a la prevención del CM.

En este trabajo cuatro modelos de evaluación de riesgo de CM fueron estudiados en profundidad, describiendo sus métodos, variables, factores de riesgo, ventajas, desventajas, limitaciones. Éstos fueron el modelo de Gail, de Claus, BRCAPRO y el Tyrer-Cuzick. Posteriormente, en una segunda etapa, el más conveniente en función de un análisis comparativo, fue desarrollado en un CDSS. De los cuatro modelos abordados, el modelo de Gail fue el seleccionado para tal desarrollo. En otras palabras, habiendo consumado estas dos fases del proyecto, los objetivos establecidos fueron cumplidos.

Aún así quedan aristas para seguir investigando y trabajando. Por un lado, tal como se mencionó en el capítulo anterior, el proceso de validación del modelo que permita aceptar, o en su defecto calibrar, el método estadístico utilizado para la evaluación de riesgo. Otros desafíos que se dejan planteados son la implementación de más de un modelo en el CDSS que permita evaluar el riesgo sobre diferentes subconjuntos poblacionales. Esto es importante para poder realizar un soporte a la decisión clínica más eficaz. También la aplicación de estándares HL7 para lograr interoperabilidad entre sistemas médicos electrónicos. Como se analizó en la sección 4.2.3, será fundamental implementar los protocolos para su integración con la historia clínica electrónica del HIBA, y demás sistemas que precisen interaccionar con la aplicación. Otro aspecto a incorporar es el manejo de bases de datos para un registro de las evaluaciones. Esto también será objetivo para la etapa posterior a este

proyecto.

Ya se han llevado a cabo encuentros con distintas partes involucradas con el objetivo de la implementación de CDSS en la historia clínica electrónica del HIBA para asistencia a los profesionales de atención primaria. La oportunidad de realizar un proyecto donde existen posibilidades reales y concretas de generar un impacto en la salud de la población, es de una gratificación inigualable.

El camino a recorrer desde la concepción de una idea como puede ser la posibilidad de predecir el riesgo a desarrollar CM, hasta su aplicación concreta en un CDSS que permita la evaluación cuantitativa, y posteriormente la adopción de medidas preventivas para resguardar la salud de las personas y mejorar así su calidad de vida, es una construcción colectiva que demanda años, esfuerzo, dedicación, recursos. Poner en perspectiva tanto conocimiento generado, desde una punta hasta la otra en este recorrido, permite contemplar un paisaje inigualable y considerar lo afortunado que se es cuando se tiene la oportunidad de ser testigo presencial de estas transformaciones en el conocimiento. Por esta razón, es necesario que los profesionales en el campo de la bioingeniería puedan aprender y entender sobre el estado del arte de los constantes desafíos en las diversas disciplinas, incluida la informática médica, para poder desarrollar estas tecnologías y ponerlas al servicio de la salud en nuestro país y en el mundo.

Apéndice A

Ejemplo de estándar FHIR en JSON para evaluación de riesgo

```
{
  "resourceType" : "RiskAssessment",
  // from Resource: id, meta, implicitRules, and
  // language
  // from DomainResource: text, contained, extension,
  // and modifierExtension
  "identifier" : { Identifier }, // Unique identifier
  // for the assessment
  "basedOn" : { Reference(Any) }, // Request fulfilled
  // by this assessment
  "parent" : { Reference(Any) }, // Part of this
  // occurrence
  "status" : "<code>", // R! registered | preliminary
  // | final | amended +
  "method" : { CodeableConcept }, // Evaluation
  // mechanism
  "code" : { CodeableConcept }, // Type of assessment
  "subject" : { Reference(Patient|Group) }, // Who/what
  // does assessment apply to?
  "context" : { Reference(Encounter|EpisodeOfCare) },
  // Where was assessment performed?
  // occurrence[x]: When was assessment made?. One of
  // these 2:
  "occurrenceDateTime" : "<dateTime>",
  "occurrencePeriod" : { Period },
```

```

"condition" : { Reference(Condition) }, // Condition
              assessed
"performer" : { Reference(Practitioner|Device) }, //
              Who did assessment?
// reason[x]: Why the assessment was necessary?. One
              of these 2:
"reasonCodeableConcept" : { CodeableConcept },
"reasonReference" : { Reference(Any) },
"basis" : [{ Reference(Any) }], // Information used
              in assessment
"prediction" : [{ // Outcome predicted
  "outcome" : { CodeableConcept }, // R! Possible
              outcome for the subject
  // probability[x]: Likelihood of specified outcome.
              One of these 2:
  "probabilityDecimal" : <decimal>,
  "probabilityRange" : { Range },
  "qualitativeRisk" : { CodeableConcept }, //
              Likelihood of specified outcome as a qualitative
              value
  "relativeRisk" : <decimal>, // Relative likelihood
              // when[x]: Timeframe or age range. One of these 2:
  "whenPeriod" : { Period },
  "whenRange" : { Range },
  "rationale" : "<string>" // Explanation of
              prediction
}],
"mitigation" : "<string>", // How to reduce risk
"comment" : "<string>" // Comments on the risk
              assessment
}

```

Apéndice B

Módulo de comunicación

```
namespace WebserviceGail
{
    public class Communication
    {
        public string agecat;
        public string agemen;
        public string ageflb;
        public string nbiops;
        public string numrel;
        public string ah;
        public string output;
        public string msg;

        public Communication()
        {
            output = "";
            msg = "";
        }

        public void readVariables(string a, string
            b, string c, string d, string e, string
            f)
        {
            agecat = a;
            agemen = b;
            ageflb = c;
```

```

        nbiops = d;
        numrel = e;
        ah = f;
    }

    public void setOutput(double prob2){

        prob2 = Math.Round(prob2, 2);
        //Image1.Visible = false;
        if ( prob2 >= 1.67)
        {
            msg = " ADVERTENCIA: La paciente es
                  de alto riesgo. Se recomienda
                  interconsulta con especialista.
                  ";
            //str =str.Replace("\n",
            Environment.NewLine);

            //Image1.Visible = true;
        }

        output = prob2.ToString() + " %";
    }
}

```

Apéndice C

Método para cálculo de riesgo

```
[WebMethod]
public double CalculateRR(string agecat,
    string agemen, string ageflb, string
    nbiops, string numrel, string ah)
{

    double rr1;
    double rr2;
    double rr3;
    double rr4;

    int age = int.Parse(agecat);
    // Debug.WriteLine(age);

    if (agemen == "7 a 11")
    {
        rr1 = 1.207;
    }
    else if (agemen == "12 a 13")
    {
        rr1 = 1.099;
    }
    else
    {
        rr1 = 1;
    }
}
```



```

// Riesgo relativos por edad y numero
de biopsias

if (age <= 50)
{
    if (nbiops == "0")
    {
        rr2 = 1;
    }
    else if (nbiops == "1")
    {
        rr2 = 1.698;
    }
    else
    {
        rr2 = 2.882;
    }
}
else
{
    //Debug.WriteLine("Mayor a 50
    *****");
    if (nbiops == "0")
    {
        rr2 = 1;
    }
    else if (nbiops == "1")
    {
        rr2 = 1.273;
    }
    else
    {
        rr2 = 1.62;
    }
}

if (ageflb == "Menor a 20")
{
    if (numrel == "0")
    {

```

```

        rr3 = 1;
    }
    else if (numrel == "1")
    {
        rr3 = 2.607;
    }
    else
    {
        rr3 = 6.798;
    }
}
else if (ageflb == "De 20 a 24")
{
    if (numrel == "0")
    {
        rr3 = 1.244;
    }
    else if (numrel == "1")
    {
        rr3 = 2.681;
    }
    else
    {
        rr3 = 5.775;
    }
}
else if (ageflb == "De 25 a 29")
{
    if (numrel == "0")
    {
        rr3 = 1.548;
    }
    else if (numrel == "1")
    {
        rr3 = 2.756;
    }
    else
    {
        rr3 = 4.907;
    }
}
}

```

```

else if (ageflb == "Sin partos")
{
    if (numrel == "0")
    {
        rr3 = 1.548;
    }
    else if (numrel == "1")
    {
        rr3 = 2.756;
    }
    else
    {
        rr3 = 4.907;
    }
}
else
{
    if (numrel == "0")
    {
        rr3 = 1.927;
    }
    else if (numrel == "1")
    {
        rr3 = 2.834;
    }
    else
    {
        rr3 = 4.169;
    }
}

if (nbiops == "0")
{
    rr4 = 1;
}
else
{
    if (ah == "No")
    {
        rr4 = 0.93;
    }
}

```

```
        else if (ah == "Si")
        {
            rr4 = 1.82;
        }
        else
        {
            rr4 = 1;
        }
    }

    //Debug.WriteLine("rr1 " + rr1);
    //Debug.WriteLine("rr2 " + rr2);
    //Debug.WriteLine("rr3 " + rr3);
    //Debug.WriteLine("rr4 " + rr4);

    return rr1 * rr2 * rr3 * rr4;
}
```

Apéndice D

Método para cálculo de probabilidad

```
// PRIMER TERMINO
term_1 = (1.0 - Math.Exp(-(h1[liminf -
    1] * Math.Exp(r[0]) + h2[liminf -
    1])) * (ageinterval[liminf - 1] -
    a))) * (h1[liminf - 1] *
    Math.Exp(r[0]) / (h1[liminf - 1] *
    Math.Exp(r[0]) + h2[liminf - 1]));

// SEGUNDO TERMINO
term_2 = (1.0 - Math.Exp(-(h1[limsup -
    1] * Math.Exp(r[1]) + h2[limsup - 1])
    * ((a + tau) - ageinterval[liminf -
    1]))) * (h1[limsup - 1] *
    Math.Exp(r[1]) / (h1[limsup - 1] *
    Math.Exp(r[1]) + h2[limsup - 1]));
// Esto es S1 * S2 (numeradores
// simplificados):
term_2 *= (Math.Exp(-(h1[liminf - 1] *
    Math.Exp(r[0]) +
    h2[liminf - 1]) *
    (ageinterval[liminf-1] - a)));

return (term_1 + term_2) * 100;
```

Bibliografía

- [1] FHIMSS Tonya J. La Lande (auth.) Eta S. Berner EdD FACMI FHIMSS (eds.) Eta S. Berner EdD, FCMI. *Clinical Decision Support Systems: Theory and Practice*. Health Informatics. Springer-Verlag New York, 2 edition, 2007.
- [2] Charles P Friedman. A “fundamental theorem” of biomedical informatics. *Journal of the American Medical Informatics Association*, 16(2):169–170, 2009.
- [3] National Breast Cancer Foundation. <http://www.nationalbreastcancer.org/breast-cancer-risk-factors>.
- [4] Eitan Amir, Orit C Freedman, Bostjan Seruga, and D Gareth Evans. Assessing women at high risk of breast cancer: a review of risk assessment models. *Journal of the National Cancer Institute*, 102(10):680–691, 2010.
- [5] Elizabeth B Claus, Neil Risch, and W Douglas Thompson. Autosomal dominant inheritance of early-onset breast cancer. implications for risk prediction. *Cancer*, 73(3):643–651, 1994.
- [6] Christi J van Asperen, MA Jonker, CE Jacobi, JEM van Diemen-Homan, E Bakker, MH Breuning, JC Van Houwelingen, and GH De Bock. Risk estimation for healthy women from breast cancer families. *Cancer Epidemiology and Prevention Biomarkers*, 13(1):87–93, 2004.
- [7] Donald A Berry, Giovanni Parmigiani, Juana Sanchez, Joellen Schildkraut, and Eric Winer. Probability of carrying a mutation of breast-ovarian cancer gene brca1 based on family history. *Journal of the National Cancer Institute*, 89(3):227–237, 1997.
- [8] Giovanni Parmigiani, Donald A Berry, and Omar Aguilar. Determining carrier probabilities for breast cancer-susceptibility genes brca1 and brca2. *The American Journal of Human Genetics*, 62(1):145–158, 1998.

- [9] Jonathan Tyrer, Stephen W Duffy, and Jack Cuzick. A breast cancer prediction model incorporating familial and personal risk factors. *Statistics in medicine*, 23(7):1111–1130, 2004.
- [10] Ethan Amir, DG Evans, Andrew Shenton, Fiona Lalloo, Anthony Moran, C Boggis, M Wilson, and Anthony Howell. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *Journal of medical genetics*, 40(11):807–814, 2003.
- [11] Catherine Meads, Ikhlaaq Ahmed, and Richard D Riley. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast cancer research and treatment*, 132(2):365–377, 2012.
- [12] Mitchell H Gail, Louise A Brinton, David P Byar, Donald K Corle, Sylvan B Green, Catherine Schairer, and John J Mulvihill. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, 81(24):1879–1886, 1989.
- [13] Joseph P Costantino, Mitchell H Gail, David Pee, Stewart Anderson, Carol K Redmond, Jacques Benichou, and H Samuel Wieand. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *Journal of the National Cancer Institute*, 91(18):1541–1548, 1999.
- [14] Elisabeth B Claus, N Risch, and W Douglas Thompson. Genetic analysis of breast cancer in the cancer and steroid hormone study. *American journal of human genetics*, 48(2):232, 1991.
- [15] Elizabeth B Claus, Neil Risch, and W Douglas Thompson. The calculation of breast cancer risk for women with a first degree family history of ovarian cancer. *Breast cancer research and treatment*, 28(2):115–120, 1993.
- [16] D Gareth R Evans and Anthony Howell. Breast cancer risk-assessment models. *Breast cancer research*, 9(5):213, 2007.
- [17] Mitchell H Gail, Joseph P Costantino, John Bryant, Robert Croyle, Laurence Freedman, Kathy Helzlsouer, and Victor Vogel. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *Journal of the National Cancer Institute*, 91(21):1829–1846, 1999.

- [18] Organización Mundial de la Salud. <http://www.who.int/mediacentre/factsheets/fs297/es/>.
- [19] Instituto Nacional del Cáncer, Ministerio de Salud de la Nación Argentina. <http://www.msal.gov.ar/inc/acerca-del-cancer/cancer-de-mama/>.
- [20] Amit X Garg, Neill KJ Adhikari, Heather McDonald, M Patricia Rosas-Arellano, PJ Devereaux, Joseph Beyene, Justina Sam, and R Brian Haynes. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama*, 293(10):1223–1238, 2005.
- [21] Lorenzo Moja, Koren H Kwag, Theodore Lytras, Lorenzo Bertizzolo, Linn Brandt, Valentina Pecoraro, Giulio Rigon, Alberto Vaona, Francesca Ruggiero, Massimo Mangia, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. *American journal of public health*, 104(12):e12–e22, 2014.
- [22] Robert C Burack and Phyllis A Gimotty. Promoting screening mammography in inner-city settings: the sustained effectiveness of computerized reminders in a randomized controlled trial. *Medical care*, 35(9):921–931, 1997.
- [23] Jon Emery, Helen Morris, Rebecca Goodchild, Thomas Fanshawe, AT Prevost, Martin Bobrow, and Ann-Louise Kinmonth. The graids trial: a cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. *British Journal of Cancer*, 97(4):486, 2007.
- [24] LE Pace and NL Keating. A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA*, 311(13):1327, 2014.
- [25] Ruth Bonita, Robert Beaglehole, and Tord Kjellström. *Epidemiologia básica*. OPS, 2008.
- [26] Breast Cancer Organization. <http://www.breastcancer.org/es/riesgo/factores>.
- [27] Michelle D Althuis, Jaclyn M Dozier, William F Anderson, Susan S Devesa, and Louise A Brinton. Global trends in breast cancer incidence and mortality 1973–1997. *International journal of epidemiology*, 34(2):405–412, 2005.

- [28] Lindsey A Torre, Freddie Bray, Rebecca L Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2):87–108, 2015.
- [29] Donna Spiegelman, Graham A Colditz, David Hunter, and Ellen Hertzmark. Validation of the gail et al. model for predicting individual breast cancer risk. *Journal of the National Cancer Institute*, 86(8):600–607, 1994.
- [30] Beverly Rockhill, Donna Spiegelman, Celia Byrne, David J Hunter, and Graham A Colditz. Validation of the gail et al. model of breast cancer risk prediction and implications for chemoprevention. *Journal of the National Cancer Institute*, 93(5):358–366, 2001.
- [31] Mitchell H Gail, Joseph P Costantino, David Pee, Melissa Bondy, Lisa Newman, Mano Selvan, Garnet L Anderson, Kathleen E Malone, Polly A Marchbanks, Wortia McCaskill-Stevens, et al. Projecting individualized absolute invasive breast cancer risk in african american women. *Journal of the National Cancer Institute*, 99(23):1782–1792, 2007.
- [32] Rayna K Matsuno, Joseph P Costantino, Regina G Ziegler, Garnet L Anderson, Huilin Li, David Pee, and Mitchell H Gail. Projecting individualized absolute invasive breast cancer risk in asian and pacific islander american women. *Journal of the National Cancer Institute*, 2011.
- [33] Adriano Decarli, Stefano Calza, Giovanna Masala, Claudia Specchia, Domenico Palli, and Mitchell H Gail. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the florence–european prospective investigation into cancer and nutrition cohort. *Journal of the National Cancer Institute*, 98(23):1686–1693, 2006.
- [34] Asociación Médica Argentina, Sociedad Argentina de Genética, Sociedad Argentina de Patología, Sociedad Argentina de Mastología, Sociedad Argentina de Radiología, Sociedad Argentina de Cancerología, Sociedad de Cirugía Plástica de Buenos Aires Asociación Argentina de Ginecología Oncológica, and Sociedad Argentina de Terapia Radiante Oncológica. Alto riesgo para cáncer de mama consenso nacional inter-sociedades. 2014.
- [35] JM Lalouel and S Yee. Pointer: a computer program for complex segregation analysis with pointers. Technical report, Tech rep, Population Genetics Laboratory, University of Hawaii, Honolulu, 1980.

- [36] Douglas F Easton, Deborah Ford, and D Timothy Bishop. Breast and ovarian cancer incidence in brca1-mutation carriers. breast cancer linkage consortium. *American journal of human genetics*, 56(1):265, 1995.
- [37] Deborah Ford, Douglas F Easton, and J Peto. Estimates of the gene frequency of brca1 and its contribution to breast and ovarian cancer incidence. *American journal of human genetics*, 57(6):1457, 1995.
- [38] David M Euhus, Kristin C Smith, Linda Robinson, Amy Stucky, Olufunmilayo I Olopade, Shelly Cummings, Judy E Garber, Anu Chittenden, Gordon B Mills, Paula Rieger, et al. Pretest prediction of brca1 or brca2 mutation by risk counselors and the computer model brcapro. *Journal of the National Cancer Institute*, 94(11):844–851, 2002.
- [39] Elizabeth B Claus, Joellen Schildkraut, Edwin S Iversen, Donald Berry, and Giovanni Parmigiani. Effect of brca1 and brca2 on the association between breast cancer risk and family history. *Journal of the National Cancer Institute*, 90(23):1824–1829, 1998.
- [40] Emanuele Mazzola, Jonathan Chipman, Su-Chun Cheng, and Giovanni Parmigiani. Recent brcapro upgrades significantly improve calibration. *Cancer Epidemiology and Prevention Biomarkers*, 23(8):1689–1695, 2014.
- [41] Hormuzd A Katki, Amanda Blackford, Sining Chen, and Giovanni Parmigiani. Multiple diseases in carrier probability estimation: accounting for surviving all cancers other than breast and ovary in brcapro. *Statistics in medicine*, 27(22):4532–4548, 2008.
- [42] Jurg Ott. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *American journal of human genetics*, 26(5):588, 1974.
- [43] Sara J Schonfeld, David Pee, Robert T Greenlee, Patricia Hartge, James V Lacey Jr, Yikyung Park, Arthur Schatzkin, Kala Visvanathan, and Ruth M Pfeiffer. Effect of changing breast cancer incidence rates on the calibration of the gail model. *Journal of Clinical Oncology*, 28(14):2411–2417, 2010.
- [44] David M Euhus. Understanding mathematical models for breast cancer risk assessment and counseling. *The breast journal*, 7(4):224–232, 2001.
- [45] V Shane Pankratz, Amy C Degnim, Ryan D Frank, Marlene H Frost, Daniel W Visscher, Robert A Vierkant, Tina J Hieken, Karthik Ghosh,

- Yaman Tarabishy, Celine M Vachon, et al. Model for individualized prediction of breast cancer risk after a benign breast biopsy. *Journal of Clinical Oncology*, 33(8):923–929, 2015.
- [46] World Wide Web Consortium. <http://www.w3c.es/Divulgacion/GuiasBreves/ServiciosWeb>.
- [47] Anne Geraci, Freny Katki, Louise McMonegal, Bennett Meyer, John Lane, Paul Wilson, Jane Radatz, Mary Yee, Hugh Porteous, and Fredrick Springsteel. *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*. IEEE Press, 1991.
- [48] Healthcare Information and Management Systems Society. *HIMSS dictionary of healthcare information technology terms, acronyms, and organizations*. HIMSS, 2013.
- [49] Health Level Seven International. <http://www.hl7.org/about/index.cfm?ref=common>.
- [50] MDCalc. <https://www.mdcalc.com/gail-model-breast-cancer-risk>.