FICHA TÉCNICA DE VARIABLES O EVENTOS

Línea de negocio: Gestión del riesgo	**Nivel:** *Riesgo Familiar*	**Patología o Evento:** *Riesgo Familiar de Cáncer de Mama*

Definición de la patología o Evento

El Riesgo Familiar de Cáncer de Mama es un constructo compuesto que identifica hogares con mayor probabilidad de que uno o más de sus miembros desarrollen esta neoplasia, como resultado de la interacción de condiciones estructurales, ambientales y sociales presentes en el entorno familiar. Este riesgo integra determinantes como el nivel socioeconómico, el contexto urbano o rural, la exposición a contaminantes ambientales y la presencia de factores demográficos y étnicos reconocidos. El índice se construye aplicando técnicas de reducción de dimensionalidad, como el Análisis de Componentes Principales (ACP), y se organiza en torno a dimensiones teóricas que incluyen vulnerabilidad socioeconómica, entorno ambiental y demografía familiar.

Fuente: Adaptado de guías de la Organización Mundial de la Salud (OMS) y revisiones científicas especializadas en epidemiología del cáncer de mama.

Objetivo	**Unidad de medición**
Cuantificar el riesgo de cáncer de mama a nivel familiar mediante un índice compuesto y estadísticamente ponderado, que permita identificar de forma temprana los hogares con mayor vulnerabilidad, orientando la priorización de intervenciones preventivas, ambientales y sociales dirigidas a reducir la incidencia y mejorar el pronóstico de la enfermedad.	De tipo porcentual (0 a 100%) representando una escala continua con media 0 y desviación estándar 1, clasificado en terciles: Bajo riesgo Riesgo medio Riesgo alto

Factores de riesgo y fuente de información

Fuentes:

- Encuesta de caracterización sociofamiliar (base estructurada)
- Registro de intervenciones psicosociales

Variables utilizadas (por dimensión):

Dimension	N_variables	PESOS
Vulnerabilidad_Economica	3	39.4%
Condiciones_Ambientales	3	60.6%

Dimension	Variable
Vulnerabilidad_Socioeconomica	estrato
Vulnerabilidad_Socioeconomica	Ingreso Economico de la Familia
Vulnerabilidad_Socioeconomica	Pertenencia Etnica
Condiciones_Ambientales	Cercanía a la vivienda relleno sanitario, rondas hídricas, canales, cementerios, humedales, terminales aéreos o terrestres
Condiciones_Ambientales	A menos de 100 metros o a una cuadra de la vivienda hay circulación de tráfico pesado
Condiciones_Ambientales	Edificaciones o vías en construcción o vías no pavimentadas a menos de una cuadra o 100 metros.

Para más detalle de los factores, ver documento anexo " *Anexo Factores de riesgo patología Cancer Mama vf*".

Forma de cálculo

1. **Selección y preparación de variables**

Descripción: La selección se hace previamente a través de las metodologicas estadísticas ACP/ACM con la finalidad de tener las varianzas explicadas por factor y por dimensión que permitan ir estructurando las ponderaciones.

Numéricas/Ordinales: Se analizaron mediante Análisis de Componentes Principales (ACP/PCA).
Categóricas/Nominales: Se analizaron mediante Análisis de Correspondencias Múltiples (ACM/MCA).

Antes de su inclusión en el análisis:
Se revisó la completitud de los datos, eliminando registros con valores faltantes en las variables clave.
Se aplicaron transformaciones necesarias, incluyendo recodificación y creación de variables dummy para los análisis requeridos.

1. **Reducción de Dimensionalidad y Asignación de Pesos**

Dimensiones con variables continuas/ordinales
Método: Análisis de Componentes Principales (ACP/PCA).
Justificación: Se evaluó la adecuación mediante el índice KMO (Kaiser-Meyer-Olkin) y la prueba de esfericidad de Bartlett.
Criterios de calidad:
KMO > 0.6: Adecuación aceptable.
Bartlett $p < 0.05$: Correlación suficiente.
Varianza explicada por el primer componente (PC1) > 40%: Suficiente capacidad de síntesis.

Dimensiones con variables categóricas
Método: Análisis de Correspondencias Múltiples (ACM/MCA).

Las varibles que salen seleccionadas son aquellas que dieron la mejor robustes a las metricas, las calificaciones para las variables y pesos actuales fueron:

Dimension	KMO	Bartlett_p	Varianza_PC1	Varianza_PC2	N_variables
Vulnerabilidad Socieconomica	0,769654	0	41,25111402	46,544376	3
Condiciones Ambientales	0,752022	0	42,87765548	30,1117654	3

Ambas dimensiones, Vulnerabilidad Socioeconómica y Condiciones Ambientales, muestran una adecuada pertinencia para el análisis mediante componentes principales, evidenciada por valores KMO superiores a 0,75, lo que indica una buena adecuación muestral para la reducción de dimensionalidad. Además, la prueba de esfericidad de Bartlett resulta significativa en ambos casos (p = 0), confirmando la existencia de correlaciones apropiadas entre las variables. El primer componente principal explica más del 41% de la varianza en ambas dimensiones, y al sumar el segundo componente, la varianza explicada se eleva a cerca del 88% para Vulnerabilidad Socioeconómica y al 73% para Condiciones Ambientales, lo que evidencia una alta capacidad de síntesis y representatividad de los datos originales. Cada dimensión está compuesta por tres variables, lo que facilita la interpretación y aplicación de los resultados en la construcción de índices compuestos robustos y comparables.

1. **Construcción de Subíndices Normalizados**

Para cada dimensión, se calculó el subíndice individual como el score (puntaje) del primer componente (PC1 para ACP, Dim1 para ACM), y se normalizó a media cero y desviación estándar uno (z-score):

$$score_{normalizado} = \frac{Score - \mu}{\sigma}$$

2. **Integración y Ponderación**

Se calculó la varianza explicada por el primer componente/dimensión en cada análisis (w_j).

Se ponderó cada subíndice normalizado por la proporción de varianza explicada en su dimensión.

Se integró el índice compuesto como la suma ponderada de los subíndices normalizados:

$$Indice_{Final} = \frac{\sum_{j=1}^{D}(score_j \; X \; w_j)}{\sum_{j=1}^{D} w_j}$$

Donde:

$score_j$: Puntaje normalizado de la dimensión j.

w_j: Varianza explicada por el primer componente de la dimensión j.

D: Número de dimensiones.

Los pesos integrados de cada variable/modalidad se calcularon multiplicando el peso interno en la dimensión por la importancia relativa de la dimensión en el índice global:

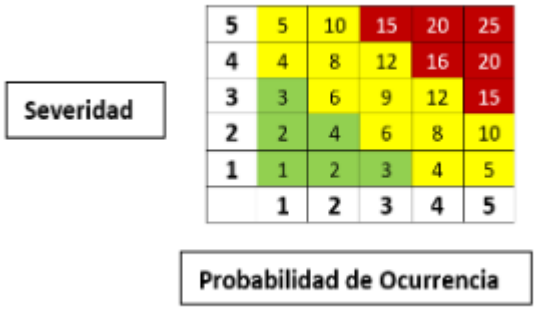
$$Peso\ Integrado\ Variable = \frac{Peso\ en\ Dimensión\ X\ w_j}{\sum_{j=1}^{D} w_j}$$

Estos pesos son los que deben emplearse para el cálculo, interpretación y pronóstico del índice en nuevas bases de datos.

Ordenamiento

Es importante considerar que, luego de la calificación registro a registro a través de la ponderación (en principio por subindices y luego con la ponderación por dimensión) se debe organizar de menor a mayor el sed de datos y este debe ser particionado en tercentiles, le primer tercentil debe ser el riesgo alto, luego riesgo medio y luego riesgo bajo.

Semaforización



La fórmula general del índice de Riesgo Familiar de cancer de mama es:

Dimensión 1: Vulnerabilidad Socioeconómica y Demográfica (peso = w_1)
Dimensión 2: Entorno Ambiental (peso = w_2)

*Cada dimensión tiene un **subíndice normalizado** ($Subíndice_1$, $Subíndice_2$), calculado según la combinación ponderada de las variables que la componen (esto ya se encuentra resuelto en el excel, usando ACP, MCA, y suma ponderada simple).*

Expresión Matemática:

$$\text{Índice de Riesgo Familiar de Cáncer de Mama} = (w_1 \times \text{Subíndice}_1) + (w_2 \times \text{Subíndice}_2)$$

$$\text{Índice de Riesgo Familiar de Cáncer de Mama} = 0.394 \times \text{Subíndice Vulnerabilidad Económica} + 0.606 \times \text{Subíndice Condiciones Ambientales}$$

$$\text{Índice}_{Familiar_i} = \frac{0.394 \cdot Z(\text{PC1}_{\text{Vulnerabilidad Económica},i}) + 0.606 \cdot Z(\text{PC1}_{\text{Condiciones Ambientales},i})}{0.394 + 0.606}$$

Como la suma es 1, la fórmula se simplifica a:

$$\boxed{\text{Índice}_{Familiar_i} = 0.394 \cdot Z(\text{PC1}_{\text{Vulnerabilidad Económica},i}) + 0.606 \cdot Z(\text{PC1}_{\text{Condiciones Ambientales},i})}$$

La fórmula general del índice de Riesgo Familiar de cáncer de mama es:

$$\text{Subíndice}_j = \sum_{i=1}^{n_j} (v_i \times p_i)$$

Donde:

- v_i: valor de la variable i en la familia (ya codificado/normalizado según el esquema del Excel)
- p_i: peso de la variable i en la dimensión j (según tu tabla)
- n_j: número de variables en la dimensión j

Bajo riesgo: Índice < 0.35
Riesgo medio: 0.35 ≤ Índice < 0.70
Alto riesgo: Índice ≥ 0.70

Algoritmo

Flujo de Trabajo Tabular: Cálculo y Semaforización del Índice de Riesgo Familiar

Etapa	Descripción/Acción	Responsable	Esperado de la herramienta
1. Análisis Estadístico Previo	Aplicar ACP (variables numéricas/ordinales) y ACM (categóricas/nominales) para reducir dimensionalidad y calcular ponderaciones de cada dimensión y variable.	Equipo estadístico	No
2. Selección de Pesos	Definir los pesos para cada dimensión (y cada opción de variable) con base en la varianza explicada u otros criterios del análisis estadístico.	Equipo estadístico	No
3. Preparación de Datos	Depurar la base de datos, recodificar variables y crear dummies según las variables seleccionadas y criterios definidos.	Usuario/soporte	No
4. **Entrega de Subíndices**	Para cada registro, entregar el subíndice normalizado (Z-score del componente principal/dimensión) por cada dimensión.	Equipo estadístico	**Sí, recibe**
5. Integración del Índice	Calcular el índice final como suma ponderada de los subíndices entregados, usando los pesos definidos previamente.	Herramienta	Sí
6. Estandarización a Escala 1-100	Transformar el índice final a escala 1-100 mediante min-max scaling.	Herramienta	Sí
7. Ordenamiento y Terciles	Ordenar los registros según el índice escalado y dividir en terciles: bajo, medio, alto riesgo.	Herramienta	Sí
8. Semaforización	Asignar color de semáforo según tercil: Verde (bajo), Amarillo (medio), Rojo (alto).	Herramienta	Sí
9. Reporte de Resultados	Exportar, visualizar y/o tabular los resultados finales, incluyendo semaforización y matriz de riesgo.	Herramienta/Usuario	Sí

Modelo (si existe)
Escribir modelo existente en el que se basa. Sino existe, dejar vacío

Periodicidad de la actualización
Semestral

Actores
SDS (Observatorio de Salud) y Subredes que interoperan en HCEU

Serie disponible (Periodo de muestra de datos)
9 meses de data para el analisis

Salidas
1. Mapa de riesgo de depresión a nivel local
2. Listado de hogares en alto riesgo de depresión
3. Apoyo a rutas de promoción y prevención del cuidado

Responsables
- Equipo de Gestión del Riesgo Familiar – SDS
- Apoyo técnico: Observatorio de Salud y Planeación Territorial

Observaciones

Este índice permite focalizar acciones intersectoriales sobre determinantes sociales de la salud para prevenir el inicio o agravamiento de condiciones de cancer de mama desde el núcleo familiar.		
Sistematización de la variable Automatización del cálculo a partir de bases de caracterización familiar, integración con tableros territoriales de riesgo en cáncer de mama y generación de alertas predictivas para equipos de promoción y prevención.		
Elaborado por: Edison Fernando Mateus Vasquez	**Aprobado por:** Equipo de proyecto Gestión de riesgo Familiar	**Fecha:** 29/05/2025