



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

INSTITUTO DE INVESTIGACIONES EN  
MATEMÁTICAS APLICADAS Y EN SISTEMAS

EFEECTO MEDIADOR DE LOS REMANENTES DE  
LIPOPROTEINAS RICAS EN TRIGLICERIDOS SOBRE  
LA RELACION ENTRE LA RESISTENCIA A LA  
INSULINA Y MORTALIDAD CARDIOVASCULAR

T E S I S

QUE PARA OBTENER EL GRADO DE:

ESPECIALISTA EN ESTADÍSTICA APLICADA

P R E S E N T A :

ARSENIO VARGAS VÁZQUEZ

TUTORES

DR. OMAR YAXMEHEN BELLO CHAVOLLA



CIUDAD UNIVERSITARIA, Cd. Mx. A 06 DE JUNIO DE 2022



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Algunas dedicatorias...*

*A mis padres, Arcenio y Cristina†.*

# Agradecimientos

- Me gustaría agradecer especialmente a mis profesores de la Especialización en Estadística Aplicada ya que con su trabajo incansable me permitieron acceder a un campo del conocimiento que a veces se considera incompatible con el ámbito médico, pero que es una herramienta indispensable para comprender a detalle el quehacer médico. Gracias a ellos y al programa he logrado consolidar aún más mi visión científica de la investigación en ciencias de la salud.
- Agradezco al Programa de la Especialización en Estadística Aplicada del IIMAS y a la Universidad Nacional Autónoma de México por la oportunidad brindada para poder continuar con mi formación académica-científica.

*Por mi raza hablará el espíritu.*

# Resumen

**Introducción:** Existe una asociación causal entre los triglicéridos (TG), las lipoproteínas ricas en TG (TRL) y los remanentes de TRL, con un mayor riesgo de morbi-mortalidad cardiovascular incluso en pacientes que reciben una terapia optimizada para reducir la concentración de lipoproteínas de baja densidad (LDL). La resistencia a la insulina es uno de los factores más importantes que regulan el ensamblaje y la secreción de TRL. Así mismo, la resistencia a la insulina a través de múltiples mecanismos participa en la fisiopatología de la enfermedad cardiovascular. Por lo tanto, en el presente trabajo se evaluó si los remanentes de colesterol actúan como un mediador parcial del riesgo que confiere la resistencia a la insulina sobre el riesgo y mortalidad cardiovascular.

**Metodología:** Utilizando los datos de National Health and Nutrition Examination Survey (NHANES), se calcularon la concentración de remanentes de colesterol (Remanentes TRL=colesterol total-colesterol HDL- colesterol LDL), usando la fórmula de Sampson y Martin para estimar el colesterol LDL. Además, para evaluar la resistencia a la insulina se calculó el índice HOMA2-IR. Se recuperaron los datos de mortalidad y se identificaron aquellos casos de mortalidad causa-específica que correspondían a mortalidad cardiovascular. Se ajustaron modelos de regresión lineal múltiple para evaluar la relación entre la resistencia a la insulina y los remanentes de colesterol y de regresión de riesgos competitivos de Fine y Gray (adaptación del modelo de Cox para riesgos competitivos) para evaluar el riesgo de mortalidad cardiovascular atribuible a HOMA2IR y a los remanentes de colesterol. Además, se realizó un análisis de mediación causal basado en modelos ordenados causalmente utilizando los coeficientes beta extraídos de los modelos de regresión de Fine y Gray.

**Resultados:** Se observó una tasa de incidencia de mortalidad cardiovascular de 1.8 (IC95 % 1.7-1.9) y 6.0 (IC95 % 5.5-6.6) eventos por cada 1000 personas-años en sujetos sin y con diabetes, respectivamente. Además, se observó una relación lineal positiva entre los remanentes de colesterol y HOMA2IR tanto en sujetos con y sin diabetes, dicha relación fue independiente del sexo, edad, IMC, etnicidad y comorbilidades,  $p$ -valor < 0.00001. Con base en los modelos de riesgos competitivos de Fine y Gray, en sujetos sin diabetes el aumento de una desviación estándar en la concentración de remanentes de colesterol se asoció con un sHR de 1.325 (IC95 % 1.060-1.658 ) y 1.289 (IC95 % 1.026-1.621) para el caso de remanentes de colesterol estimado por la fórmula de Sampson y Martin, respectivamente. En cambio, en sujetos con diabetes los niveles

de remanentes de colesterol  $Rem_S \geq 40.0$  /  $Rem_M \geq 38.0$   $mg/dL$  se asociaron con un sHR 1.285 (IC95 % 1.016-1.626) y 1.295 (IC95 % 1.011-1.659), respectivamente. Además, un  $HOMA2IR \geq 2.5$  o  $\geq 4.9$  se asoció con mayor incidencia relativa de mortalidad cardiovascular en sujetos sin diabetes, pero no en sujetos con diabetes. Por último, con base en los análisis de mediación causal se estimó que alrededor del 18 % del efecto total de la resistencia a la insulina sobre mortalidad cardiovascular está parcialmente mediado por los remanentes de colesterol en sujetos sin diabetes. En cambio, en sujetos con diabetes los remanentes de colesterol no median la relación entre la resistencia a la insulina y mortalidad cardiovascular.

**Conclusiones:** El efecto (total) que tiene la resistencia a la insulina sobre la mortalidad cardiovascular está parcialmente mediado por los remanentes de colesterol en sujetos sin diabetes, no así en sujetos con diabetes. Debido a que este abordaje considera que los mediadores están causalmente ordenados, la resistencia a la insulina condiciona un aumento en los niveles de remanentes de colesterol y a su vez esto aumenta el riesgo de mortalidad cardiovascular en sujetos sin diabetes, por lo que el mecanismo tiene plausibilidad biológica. La aplicación de modelos de mediación causal en el contexto del análisis de supervivencia es una técnica útil cuando además de informar el *por qué*, se busca el *cómo* de los resultados. Sin embargo, es necesario el desarrollo de un conjunto de criterios que permitan evaluar acorde al tipo de evento a estudiar, así como el método utilizado, la validez de los modelos de mediación causal.

# Índice general

<b>Agradecimientos</b>	<b>II</b>
<b>Resumen</b>	<b>III</b>
<b>1. Introducción</b>	<b>1</b>
1. Remanentes de lipoproteínas ricas en triglicéridos (TRL) . . . . .	1
1.1. Resistencia a la insulina y TRL . . . . .	2
1.2. Riesgo cardiovascular y remanentes TRL . . . . .	3
1.3. Riesgo cardiovascular y resistencia a la insulina . . . . .	4
1.4. Modelos de mediación causal para evaluar riesgo cardiovascular . . . . .	5
2. Planteamiento del problema . . . . .	6
3. Objetivos . . . . .	7
4. Hipótesis . . . . .	7
<b>2. Metodología</b>	<b>8</b>
1. Población de estudio; National Health and Nutrition Examination Survey (NHANES) . . . . .	8
1.1. Evaluaciones bioquímicas y antropométricas . . . . .	10
2. Análisis estadístico . . . . .	11
2.1. Modelo de regresión lineal múltiple . . . . .	11
2.1.1. Ajuste del modelo RLM para evaluar la relación entre los remanentes de colesterol y la resistencia a la insulina . . . . .	15
2.2. Modelo de riesgos competitivos . . . . .	16
2.2.1. Especificación del modelo de riesgos competitivos . . . . .	18
2.2.2. Funciones de riesgo causa-específica $\lambda_j(t)$ ; modelo de riesgos proporcionales de Cox . . . . .	19
2.2.3. El modelo de Fine y Gray . . . . .	22
2.2.4. Implementación del modelo de Fine y Gray . . . . .	25
2.2.5. Evaluación del modelo de Fine y Gray . . . . .	26
2.3. Modelo de Mediación Causal (AMC) . . . . .	27
2.3.1. Análisis de mediación causal en análisis de supervivencia . . . . .	29
2.3.2. Análisis de mediación causal en modelos de riesgos proporcionales . . . . .	31
2.3.3. Implementación del modelo de mediación causal: mediadores causalmente ordenados . . . . .	33

<b>3. Resultados</b>	<b>38</b>
1. Análisis descriptivo: características clínicas, bioquímicas, antropométricas y mortalidad. . . . .	38
2. Modelos de regresión lineal múltiple . . . . .	47
2.1. Modelos de regresión lineal múltiple en sujetos sin diabetes . . . . .	47
2.1.1. Modelo 1 de RLM para remanentes de colesterol (Sampson) en sujetos sin diabetes . . . . .	47
2.1.2. Modelo 2 de RLM para remanentes de colesterol (Martin) en sujetos sin diabetes . . . . .	52
2.2. Modelos de regresión lineal múltiple en sujetos con diabetes . . . . .	56
2.2.1. Modelo 3 de RLM para remanentes de colesterol (Sampson) en sujetos con diabetes . . . . .	56
2.2.2. Modelo 4 de RLM para remanentes de colesterol (Martin) en sujetos con diabetes . . . . .	59
3. Modelos de riesgos competitivos de Fine y Gray . . . . .	63
3.1. Modelos de riesgos competitivos de Fine y Gray en sujetos sin diabetes . . . . .	64
3.1.1. Modelos 1 de riesgos competitivos de Fine y Gray en sujetos sin diabetes . . . . .	64
3.1.2. Modelos 2 de riesgos competitivos de Fine y Gray en sujetos sin diabetes . . . . .	68
3.2. Modelos de riesgos competitivos de Fine y Gray en sujetos con diabetes . . . . .	73
3.2.1. Modelos 3 de riesgos competitivos de Fine y Gray en sujetos con diabetes . . . . .	73
3.2.2. Modelo 4 de riesgos competitivos de Fine y Gray en sujetos con diabetes . . . . .	78
4. Análisis de Mediación Causal . . . . .	82
4.1. Análisis de Mediación Causal en sujetos sin diabetes . . . . .	82
4.2. Análisis de Mediación Causal en sujetos con diabetes . . . . .	83
<b>4. Discusión y Conclusiones</b>	<b>86</b>

# Capítulo 1

## Introducción

### 1. Remanentes de lipoproteínas ricas en triglicéridos (TRL)

Los remanentes de lipoproteínas ricas en triglicéridos (TRL), también conocidas simplemente como remanentes, son los productos de partículas de lipoproteínas ricas en triglicéridos (TG), que han sido metabolizadas principalmente por lipasas para producir partículas enriquecidas con colesterol. Los remanentes de TRL incluyen restos de quilomicrones, lipoproteínas de muy baja densidad (VLDL) y lipoproteínas de densidad intermedia (IDL) [1]. Hormonas, nutrientes, señales neuronales y, múltiples enzimas y proteínas regulan el ensamblaje y secreción de los quilomicrones en el intestino delgado[2]. En cambio, para las VLDL los efectos crónicos del exceso de nutrientes y la resistencia a la insulina son los factores más importantes, incrementando la producción hepática de VLDL[3].

La depuración de las TRL incluye dos procesos interrelacionados; la lipólisis de TG por las lipoproteínas lipasa (LpL) y la depuración hepática de los remanentes. El tamaño de los TRL disminuye durante la lipólisis. Durante este proceso, el contenido de TG decrece y el contenido de colesterol aumenta. A niveles óptimos de TG (< 100 mg/dL), la lipólisis resulta eficiente y la acumulación de TRL es limitada, predomi-

nando VLDL pequeñas e IDL. Sin embargo, los niveles elevados de TG ( $\geq 260\text{mg/dL}$ ) condicionan el aumento en la secreción y la alteración de la lipólisis que dan como resultado la acumulación sustancial de TRL, principalmente restos de quilomicrones y VLDL. Los TRL son depurados directamente en el hígado o convertidos a LDL, en la mayoría de las personas las partículas de TRL son muy heterogéneas, por lo que especificar qué características se asocian a riesgo cardiovascular resulta desafiante [4].

### 1.1. Resistencia a la insulina y TRL

Los componentes de la dislipidemia que ocurren cuando se presenta resistencia a la insulina son; 1) hipertrigliceridemia, 2) disminución de los niveles de HDL y 3) cambios en la composición de las lipoproteínas (TRL y LDL)[5]. La hiperinsulinemia y la obesidad central, que típicamente acompaña a la resistencia a la insulina, condicionan la sobreproducción de VLDL debido a:

- **i)** Aumento del suministro de ácidos grasos liberados por los adipocitos resistentes a la insulina que participan como sustrato para la síntesis de VLDL.
- **ii)** Aumento del suministro de TG-ácidos grasos al hígado debido a la reducción de la lipólisis periférica de quilomicrones y VLDL-TG. El aumento del suministro de TG al hígado inhibe la degradación de la apolipoproteína B (proteína estructural de las VLDL) incrementando el ensamblaje y secreción de VLDL. Además, la reducción de la lipólisis periférica de quilomicrones y VLDL-TG es debido a la reducción de los niveles de LpL y un aumento en la expresión de ApoC-III (inhibidor de la LpL), que condicionan una disminución en el aclaramiento de las TRL.
- **iii)** Aumento de la lipogénesis de novo en el hígado [6].

## 1.2. Riesgo cardiovascular y remanentes TRL

Recientemente, estudios genéticos en humanos, junto con una gran cantidad de resultados de ensayos clínicos, preclínicos y epidemiológicos, han demostrado una asociación causal entre los TG, las TRL y los remanentes de TRL, con un mayor riesgo de infarto de miocardio, accidente cerebrovascular isquémico y estenosis de la válvula aórtica [4]. Además, las TRL y sus remanentes pueden contribuir significativamente al riesgo cardiovascular residual en pacientes que reciben una terapia optimizada para reducir la concentración de lipoproteínas de baja densidad (LDL) [7].

La explicación del efecto causal sobre mortalidad cardiovascular de los niveles elevados del colesterol de los remanentes TRL probablemente sea simple y directa. Es decir, que tanto el colesterol de los remanentes de TRL como el colesterol-LDL ingresa y queda atrapado en la íntima media de las paredes de las arterias causando la acumulación de colesterol en la pared arterial, aterosclerosis y, en última instancia eventos cardiovasculares [8, 9]. Sin embargo, se debe considerar que el tamaño de las lipoproteínas desempeña un rol importante en su capacidad para depositarse en las paredes de los vasos sanguíneos [8, 10]. Estudios en humanos y en animales han mostrado que los quilomicrones y las VLDL grandes son incapaces de penetrar en la pared arterial, pero sus remanentes son capaces de entrar y quedar atrapados en el tejido conectivo de la pared arterial hasta eventualmente acumularse [11]. Además, estudios *in vitro* han demostrado que los remanentes de TRL a diferencia de las LDL pueden no necesitar oxidarse para que los macrófagos los absorban y provoquen la formación de células espumosas y aterosclerosis [12].

Los remanentes de TRL pueden estar involucrados en el desarrollo de aterosclerosis y de cardiopatía isquémica por otras vías además de la acumulación de colesterol en la pared arterial. Diversos estudios han demostrado que los remanentes están relacionados con mecanismos subyacentes a la disfunción endotelial; vasodilatación alterada, respuesta proinflamatoria y que tienen un papel importante en la ruptura de la placa

ateroesclerótica y la formación de trombos [13, 14, 15]. Estudios de aleatorización Mendeliana multidireccional han mostrado que niveles elevados de remanentes de TRL pueden no causar aterosclerosis exactamente de la misma forma que el colesterol contenido en las LDL. En este caso, los niveles de remanentes de TRL se asocian causalmente tanto con inflamación de bajo grado como con cardiopatía isquémica; en cambio, los niveles de colesterol LDL se asocian causalmente con cardiopatía isquémica, pero no con inflamación de bajo grado. Por lo tanto, una concentración elevada de colesterol LDL causa aterosclerosis sin un componente inflamatorio importante, mientras que el componente inflamatorio de la aterosclerosis puede estar mediado por los niveles de remanentes colesterol [16].

### 1.3. Riesgo cardiovascular y resistencia a la insulina

La resistencia a la insulina es un término que describe el fenómeno en el que la insulina no ejerce sus efectos normales en los tejidos diana sensibles a la misma, como el músculo esquelético, el tejido adiposo, el hígado y el páncreas [17].

La resistencia a la insulina, medida mediante con el clamp euglicémico hiperinsulinémico (estándar de oro), se asocia con aterosclerosis asintomática en personas sin diabetes mellitus [18, 19]. Además, los niveles de insulina en ayunas, un marcador de resistencia a la insulina, también se han asociado con eventos cardiovasculares en personas sin diabetes, independientemente de otros factores de riesgo [20, 21].

Los índices subrogados de resistencia a la insulina como  $HOMA-IR$  son mejores predictores de eventos cardiovasculares que los niveles aislados de glucosa o insulina en ayuno en adultos sin diabetes [22]. Sin embargo, la proporción de enfermedades cardiovasculares que puede atribuirse a la resistencia a la insulina aún no está completamente descrita. Para responder a esta pregunta, se ha aplicado el modelo de Arquímedes para estimar la proporción de eventos de infarto de miocardio que se evitarían manteniendo la resistencia a la insulina y los factores de riesgo cardiovascular en niveles normales. Utilizando las Encuesta Nacionales de Salud en los EE. UU., se

reportó que la prevención de la resistencia a la insulina podría prevenir aproximadamente el 42% de los infartos de miocardio en adultos jóvenes durante un período de seguimiento simulado de 60 años [23].

Además, el hallazgo más temprano en la fisiopatología de la aterosclerosis es la alteración de la función endotelial, que está estrechamente relacionada con la resistencia a la insulina. La disminución de la síntesis y liberación de óxido nítrico endotelial relacionado a resistencia a la insulina podría contribuir a un mayor riesgo de aterosclerosis en estos pacientes [24, 25]. Sin embargo, no está claro si la resistencia a la insulina endotelial afecta la progresión de las placas necróticas, que son lesiones clínicamente relevantes en humanos [26]. Los estudios en humanos respaldan el papel del estrés del retículo endoplásmico a largo plazo en la apoptosis de los macrófagos y la necrosis de la placa, que a su vez puede conducir a la ruptura de la placa, la formación de trombos y, en consecuencia, a eventos cardiovasculares [27, 28].

#### **1.4. Modelos de mediación causal para evaluar riesgo cardiovascular**

A pesar de que diversos estudios abordan la relación entre los diferentes factores de riesgo y la enfermedad cardiovascular en poblaciones específicas, se desconoce la contribución relativa de estos factores al riesgo cardiovascular. En la investigación epidemiológica, la comprensión de las vías causales desde la exposición hasta un resultado de interés puede evaluarse utilizando técnicas de análisis de mediación. El análisis de mediación ofrece una herramienta útil para evaluar la magnitud de las diferentes vías (mediadores) que conducen a un resultado en específico [55, 30].

Brevemente, este tipo de análisis proporciona una herramienta para deconstruir el efecto total de una exposición determinada sobre un resultado en un efecto natural directo y un efecto indirecto. El efecto total estima el riesgo de eventos por todas las causas. La interpretación de estos efectos es la siguiente: el efecto indirecto es la

diferencia (relativa) de eventos atribuibles a la mediación a través del factor de riesgo, mientras que el efecto directo natural es la diferencia (relativa) de eventos atribuibles a un camino directo. Finalmente, el efecto total es la diferencia (relativa) en eventos entre quienes están expuestos a un factor de riesgo en comparación con quienes no presentan el factor de riesgo, lo que equivale a la suma del efecto directo natural y el efecto indirecto [55, 31].

Algunos estudios han evaluado el efecto mediador de ciertos factores sobre la enfermedad y/o mortalidad cardiovascular. Por ejemplo, se ha logrado determinar que los factores sociales y conductuales de los adultos tienen un efecto mediador considerable en el origen social de la mortalidad por eventos cardiovasculares y otras causas en los primeros años de vida. Además, algunos estudios han reportado que una cantidad sustancial del aumento del riesgo cardiovascular y de la mortalidad por cualquier causa debida a la diabetes tipo 2 no puede explicarse por los factores de riesgo vascular tradicionales. Lo que sugiere el desarrollo de investigación futura centrada en identificar vías no clásicas que puedan explicar mejor el aumento del riesgo cardiovascular y de mortalidad causado por la diabetes Tipo 2 o por otra causa [55, 32].

## 2. Planteamiento del problema

Se ha establecido que la resistencia a la insulina, así como las TRL y sus remanentes son predictores importantes de riesgo y mortalidad cardiovascular, incluso en pacientes con niveles óptimos de colesterol LDL. Además, la resistencia a la insulina es uno de los factores más importantes que regulan el ensamblaje, secreción y depuración de las TRL y sus remanentes. Por lo tanto, es de interés explorar si los remanentes de colesterol pueden actuar como un mediador del riesgo que confiere la resistencia a la insulina sobre el riesgo y mortalidad cardiovascular. Con este fin, resulta de utilidad realizar una aproximación epidemiológica conocida como análisis de mediación causal, el cual nos permitirá deconstruir el efecto total de la resistencia a

la insulina sobre el riesgo y/o mortalidad cardiovascular, en un efecto natural directo explicado por los mecanismos fisiopatológicos atribuibles a la resistencia a la insulina y un efecto indirecto atribuible a las TRL y sus remanentes. Por lo tanto, se establecerá un porcentaje de mediación atribuible a las TRL y sus remanentes de un efecto total dado por la resistencia a la insulina.

### 3. Objetivos

- **Objetivo principal:** Determinar la proporción de mediación causal atribuible a las TRL y sus remanentes del efecto total de la resistencia a la insulina sobre la mortalidad cardiovascular.
  - **Objetivo específico 1:** Identificar los principales predictores de la concentración de TRL y sus remanentes.
  - **Objetivo específico 2:** Estimar el efecto indirecto atribuible a las TRL y sus remanentes del efecto total de la resistencia a la insulina sobre la mortalidad cardiovascular en sujetos sin diabetes.
  - **Objetivo específico 3:** Estimar el efecto indirecto atribuible a las TRL y sus remanentes del efecto total de la resistencia a la insulina sobre la mortalidad cardiovascular en sujetos con diabetes.

### 4. Hipótesis

El efecto natural total de la resistencia a la insulina sobre la mortalidad cardiovascular está parcialmente mediado (efecto indirecto) por las TRL y sus remanentes. Basado en las consideraciones planteadas, el modelo de mediación causal estará compuesto de un estimador de resistencia a la insulina (HOMA2IR), una estimación de TRL y sus remanentes (fórmula de Martín y Sampson) y mortalidad cardiovascular (eventos).

# Capítulo 2

## Metodología

### 1. Población de estudio; National Health and Nutrition Examination Survey (NHANES)

La Encuesta Nacional de de Salud y Nutrición (NHANES) es un programa de estudios diseñado para evaluar el estado de salud y nutrición de adultos y niños en los Estados Unidos. NHANES es un programa principal del Centro Nacional de Estadísticas de Salud (NCHS). NCHS es parte de los Centros para el Control y la Prevención de Enfermedades (CDC) y tiene la responsabilidad de producir estadísticas vitales y de salud para los Estados Unidos [33].

Se seleccionó la Tercera Encuesta Nacional de de Salud y Nutrición, NHANES III, una muestra probabilística nacional de 39,695 personas de 2 meses o más. Se llevó a cabo entre 1988 y 1994 en dos fases. La Fase 1 (1988-1991) y la Fase 2 (1991-1994) fueron muestras representativas a nivel nacional. En la muestra de seis años, se entrevistó a 33 994 personas de la muestra y se examinó a 30 818, lo que dio como resultado una tasa de respuesta examinada general del 78 por ciento. Los niños de dos meses a cinco años, las personas de 60 años o más, las personas mexicoamericanas y las personas de raza negra no hispana fueron muestreadas a tasas sustancialmente más altas que sus proporciones en la población general [33].

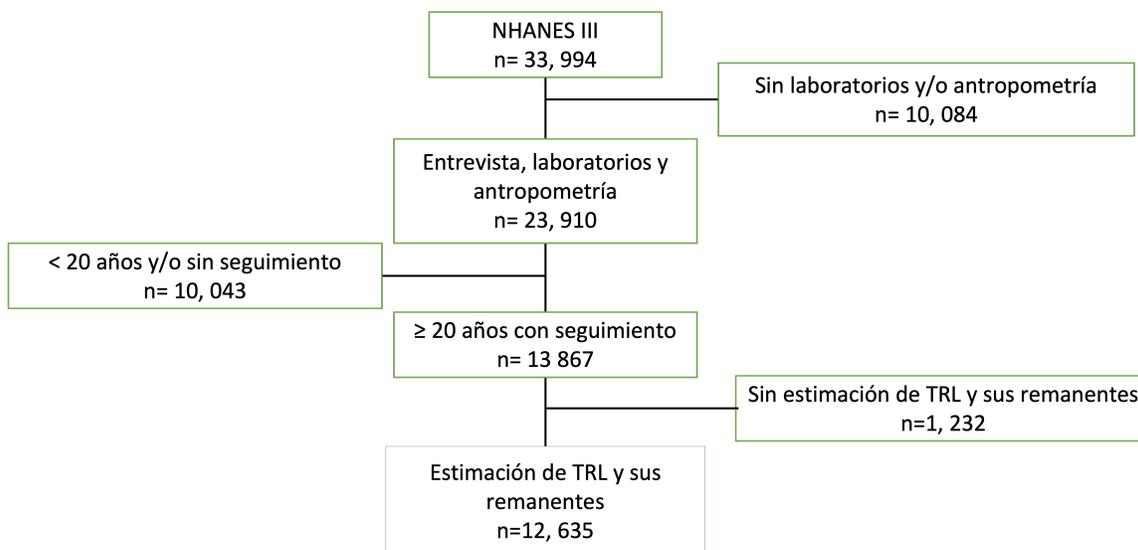


Figura 2.1: Flujograma de la selección de la muestra a partir de NHANES III

Como se observa en la Figura 2.1, del total de sujetos incluidos en NHANES III se eliminaron aquellos menores de 20 años de edad, sin datos de seguimiento, laboratorio y/o antropometría. De los 13, 867 sujetos mayores de 20 años de edad y con datos de seguimiento, laboratorio y/o antropometría, se eliminaron del análisis aquellos en los que no fue posible calcular TRL y sus remanentes debido a que no tenían reportado algún parámetro de laboratorio necesario para la estimación. Finalmente, se analizaron a 12, 635 sujetos.

De las aproximadamente 4000 variables incluidas en NHANES III, sólo se seleccionaron aquellas variables que recuperaran información acerca de los antecedentes personales patológicos (diabetes, hipertensión, insuficiencia cardiaca, cardiopatía isquémica, enfermedad cerebrovascular, etc.), sexo, edad, etnicidad (afroamericanos, mexicanoamericanos, caucásicos y otros), tiempo de seguimiento (tiempo en meses desde la fecha de entrevista hasta confirmación del estado de mortalidad en 2018), mortalidad por cualquier causa y mortalidad específica (enfermedades del corazón, enfermedad cerebrovascular, neoplasias malignas, accidentes, diabetes, influenza o neumonía, Alzheimer, enfermedad pulmonar obstructiva crónica, otras causas). La variable

mortalidad cardiovascular, se definió como la combinación de mortalidad asociada a enfermedades del corazón y/o mortalidad asociada a enfermedad cerebrovascular. El diagnóstico de diabetes se definió como un diagnóstico previo, hemoglobina glucosilada  $\geq 6.5\%$  y/o glucosa en ayuno  $\geq 126\text{mg/dL}$ .

### 1.1. Evaluaciones bioquímicas y antropométricas

Los procedimientos para la obtención de las variables bioquímicas y antropométricas ha sido descrito previamente [34]. Se recuperaron las siguientes variables; glucosa, hemoglobina glucosilada (HbA1C), péptido C, insulina, colesterol total, colesterol HDL, triglicéridos, apolipoproteína B, peso, talla y circunferencia de cintura.

El índice de masa corporal (IMC) se calculó como el peso en kilogramos dividido entre el cuadrado de la talla en metros. El índice cintura-estatura se calculó como la circunferencia de cintura en centímetros dividido entre la estatura en centímetros. Además, se calcularon utilizando los niveles de glucosa y péptido C en ayuno los índices de resistencia a la insulina (HOMA2IR), función de célula beta (HOMA2B) y sensibilidad a la insulina (HOMA2S) con la calculadora disponible en <https://www.dtu.ox.ac.uk/homacalculator/>. Las TRL y sus remanentes se calcularon con la siguiente fórmula:

$$\text{Remanentes} = \text{Colesterol Total} - \text{Colesterol HDL} - \text{Colesterol LDL}, \quad (2.1)$$

El colesterol LDL se estimó a partir de dos fórmulas. La fórmula de Martín ( $LDL_M$ ) que utiliza un factor ajustable de la relación TG:VLDL-C. Utilizando la mediana de la relación TG:VLDL-C se especifica el factor ( $x_i$ ) acorde a los niveles de TG y de colesterol no HDL [35]. Tal como se describe en la siguiente fórmula:

$$LDL_M = \text{Colesterol Total} - \text{Colesterol HDL} - \frac{TG}{x_i} \quad (2.2)$$

En cambio la fórmula de Sampson ( $LDL_S$ ) desarrollada para mejorar la estimación

del colesterol LDL en pacientes con hipertrigliceridemia o incluso con niveles bajos de triglicéridos [36], se describe a continuación:

$$LDL_S = \frac{CT}{0.948} - \frac{HDL - C}{0.971} - \frac{TG}{8.59} + \frac{(TG * noHDL - C)}{2140} - \frac{TG^2}{16100} - 9.44 \quad (2.3)$$

Donde, HDL-C indica el colesterol HDL y noHDL-C el colesterol no HDL que se estimó al restar el colesterol total menos el colesterol HDL.

## 2. Análisis estadístico

Para evaluar el efecto mediador de los TRL y sus remanentes sobre la relación entre la resistencia a la insulina y la mortalidad cardiovascular, se ajustaron los siguientes modelos:

- Regresión lineal múltiple.
- Modelo de riesgos competitivos de Fine and Grey
- Modelo de mediación con mediadores causalmente ordenados.

A continuación se describirá la teoría detrás de cada uno de los modelos, así como las particularidades de su implementación y ajuste.

### 2.1. Modelo de regresión lineal múltiple

Un modelo en el que la variable de respuesta ( $Y$ ) depende de dos o más variables predictoras ( $X_1, X_2, \dots, X_p$ ) se le conoce como regresión múltiple. Si la relación entre los parámetros es lineal, entonces se le conoce como *modelo de regresión lineal múltiple* (RLM) [37, 38]. El modelo RLM busca formular la relación entre  $Y$  y  $X_1, X_2, \dots, X_p$  como un modelo lineal de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon_i, \quad (2.4)$$

donde  $X_1, X_2, \dots, X_p$  son los valores de entrada de las variables predictoras para el  $i$ -ésimo observación y  $Y_i$  la respuesta. Además,  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  son términos constantes contenidos en un vector  $\hat{\beta}$  referidos como coeficientes de la regresión y  $\epsilon$  corresponde al error que se asumen que son i.i.d con distribución normal de media 0 y varianza  $\sigma^2$  desconocida. Particularmente, el modelo de RLM es una generalización de la regresión lineal simple (RLS), es decir, el modelo RLS podría expresarse como un caso especial del modelo RLM cuando el número de predictores ( $p$ ) es  $p = 1$  [39].

El modelo RLM realiza una predicción lineal de la relación de la variable aleatoria dependiente  $Y_i$  sobre una serie de predictores  $X_1, X_2, \dots, X_n$ , variables independientes. Para esto, el modelo RLM estima los parámetros del vector  $\hat{\beta}$  mediante el método de mínimos cuadrados ordinarios (MCO). El método MCO busca minimizar la suma del cuadrado de los errores que tiene como solución un sistema de ecuaciones normales. Usando el método de mínimos cuadrados, se busca minimizar

$$S(\beta) = \sum_{i=1}^n \epsilon_i^2 = \epsilon' \epsilon = (\mathbf{y} - \mathbf{x}\beta)'(\mathbf{y} - \mathbf{x}\beta) \quad (2.5)$$

$$S(\beta) = \mathbf{y}'\mathbf{y} - \beta'\mathbf{x}'\mathbf{y} + \mathbf{y}'\mathbf{x}\beta + \beta'\mathbf{x}'\mathbf{x}\beta \quad (2.6)$$

$$S(\beta) = \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{x}'\mathbf{y} + \beta'\mathbf{x}'\mathbf{x}\beta \quad (2.7)$$

Considerando que  $\beta'\mathbf{x}'\mathbf{y}$  es un escalar, entonces  $\beta'\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}\beta$ . Entonces se deriva con respecto a  $\beta$  y se iguala a 0. Por lo tanto, se satisface que

$$\frac{\partial S}{\partial \beta} \Big|_{\hat{\beta}} = -2\mathbf{x}'\mathbf{y} + 2\mathbf{x}'\mathbf{x}\hat{\beta} = 0 \quad (2.8)$$

$$\mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{x}\hat{\beta} \quad (2.9)$$

Para solucionar el sistema de ecuaciones se multiplica por ambos lados por  $(\mathbf{x}'\mathbf{x})^{-1}$ . Por lo tanto, los estimadores por mínimo cuadrados está dado por

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \quad (2.10)$$

Es importante mencionar que el estimador por mínimos cuadrados  $\hat{\boldsymbol{\beta}}$  siempre está bien definido si  $(\mathbf{x}'\mathbf{x})^{-1}$  existe. Por ende, la inversa existe siempre que las variables predictoras sean linealmente independientes.

Entonces, el vector de los valores ajustados para  $\hat{y}_i$  estará dado por

$$\hat{\mathbf{y}} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} = \mathbf{H}\mathbf{y} \quad (2.11)$$

Donde  $\mathbf{H} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$ , de tal forma que los errores están dados por

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \quad (2.12)$$

Los estimadores de MCO tienen las siguientes propiedades:

- El estimador  $\hat{\beta}_j$  y el estimador  $\hat{\sigma}^2$  son estimadores insesgados de  $\beta_j$  y  $\sigma^2$ , respectivamente. Para todos los estimadores lineales los estimadores MCO tienen varianza mínima por lo que se les considera los *mejores estimadores lineales insesgados (MELI)*.
- El estimador  $\hat{\beta}_j$   $j = 0, 1, \dots, p$  tiene distribución normal con media  $\beta_j$  y varianza  $\sigma^2 c_{ij}$ , donde  $c_{ij}$  es el  $j$ -ésimo elemento del inverso de la matriz de covarianza de los estimadores  $\hat{\beta}_j$  y  $\hat{\beta}_i$ .
- El estimador  $W = SCE/\sigma^2$  tiene distribución  $\chi^2$  con  $n - p - 1$  grados de libertad, donde  $\hat{\boldsymbol{\beta}}$  y  $\hat{\sigma}^2$  se distribuyen independiente el uno del otro.

Las propiedades previas de los estimadores MCO son de utilidad cuando se busca hacer inferencias sobre las  $\hat{\beta}_j$  o sobre  $\hat{\sigma}^2$ . Sin embargo, sólo si se cumplen los supuestos del modelo RLM es posible hacer inferencias. Dado que se puede pensar al modelo de RLM como un caso general del modelo de regresión lineal simple, éste se basa en 10 supuestos. El número de supuestos puede variar dependiendo la bibliografía ya que algunos supuestos se dan por obvios, se fusionan o no se enumeran [39].

- **Correcta especificación del modelo:** este supuesto hace referencia a la inclusión de variables independientes que aporten información relevante que permita predecir a la variable respuesta. Por lo tanto, el reto implica no obtener un modelo tan complejo que sea difícil de interpretar o tan simple que no describa al fenómeno estudiado.
- **Covarianza entre las variables regresoras y el error:** Es esperado que los datos han sido obtenidos de algún experimento o fueron proporcionados por alguien más, por lo tanto, son fijos, aunque describan un fenómeno aleatorio. Entonces, si la distribución de  $y$  dada  $x$  se ajusta a una distribución normal y la distribución de  $x$  no depende de ninguno de los parámetros desconocidos, los estimadores de mínimos cuadrados siguen siendo válidos.
- **Supuestos sobre la forma del modelo:** El modelo RLM que relaciona una variable  $Y$  con los predictores  $X_1, X_2, \dots, X_p$  es lineal para los estimadores de regresión  $\beta_1, \beta_2, \dots, \beta_n$ . A éste supuesto se le conoce como supuesto de *linealidad*.
- **Supuestos sobre los errores del modelo:** Se asume que los errores del modelo son variables aleatorias normales idénticas e independientemente distribuidas (i.i.d.) con media 0 y varianza  $\sigma^2$ . Entonces:
  - Los errores  $\epsilon_i$  tienen distribución normal con media cero. A este supuesto se le conoce como *distribución normal de los errores*.
  - Los errores  $\epsilon_i$  tienen varianza desconocida  $\sigma^2$  constante a lo largo de toda la predicción de la variable dependiente. Esto se conoce como el supuesto de *varianza constante u homoscedasticidad*.
  - Los errores son independientes con covarianzas por pares igual a cero. A este supuesto se le conoce como de *independencia de los errores*.
- **Multicolinealidad:** indica la existencia de una relación lineal entre dos o más variables independientes. Este supuesto obedece a que la matriz debe ser de rango completo, es decir, los predictores dentro del modelo no se componen de

una combinación lineal de otros predictores. Cuando existe multicolinealidad perfecta, los coeficientes de la regresión no estarán bien definidos y su varianza será infinita. En caso de multicolinealidad no perfecta, los coeficientes tendrán una varianza muy grande. Hay una relación lineal perfecta entre las variables independientes si existen constantes no todas cero  $\lambda_1, \dots, \lambda_p$  y, por lo tanto;

$$\lambda_1 x_{i1} + \dots + \lambda_p x_{ip} = 0 \quad (2.13)$$

En cambio existe una relación lineal no perfecta cuando

$$\lambda_1 x_{i1} + \dots + \lambda_p x_{ip} + \lambda_{p+1} = 0 \quad (2.14)$$

### 2.1.1. Ajuste del modelo RLM para evaluar la relación entre los remanentes de colesterol y la resistencia a la insulina

1. Se realizó un ajuste de un modelo RLM para evaluar la relación de Remanentes de colesterol como variable respuesta y como variables independientes; HOMA2IR, edad, sexo, etnicidad, IMC, tabaquismo, colesterol LDL y número de comorbilidades. Antes del ajuste del modelo, el total de la muestra se segmentó acorde al diagnóstico de diabetes. Por lo tanto, se ajustaron modelos de RLM en pacientes con y sin diabetes por separado. Se validaron los supuestos del modelo usando métodos gráficos y las pruebas estadísticas específicas para cada supuesto, dado el caso. En caso de no cumplirse alguno de los supuestos, se optó por transformar a las variables para mejorar el ajuste del modelo.
2. Los modelos fueron evaluados para los supuestos del modelo RLM:
  - **Linealidad:** Se realizaron pruebas de falta de ajuste utilizando términos cuadráticos en los modelos RLM para cada una de las variables del modelo. Falta de ajuste se definió como una falta de rechazo de la hipótesis nula de  $\beta_i = 0$  con un nivel de confianza del 95 %, por lo tanto, con un nivel de significancia establecido en  $\alpha = 0.05$ .

- **Homoscedasticidad de varianza:** se realizó la prueba de Breusch-Pagan, así como la observación de los gráficos de los residuos para evaluar independencia de los errores.
  - **Matriz de diseño de rango completo:** Se evaluó la colinealidad de las variables en el modelo, para validar que la colinealidad no influyó en el modelo y que ésta no está presente en el modelo final, se estimó el factor de inflación de la varianza (VIF). Donde un valor de  $VIF < 5$  indicó ausencia de colinealidad significativa en el modelo..
  - **Normalidad conjunta de los errores:** Se exploró gráficamente los residuos y, posteriormente se probó su normalidad con la prueba de Anderson-Darling y Jarque-Bera en caso de no independencia de los errores.
  - **No autocorrelación de los errores:** Se realizó la prueba de Breusch-Godfrey para evaluar la no autocorrelación de los errores.
3. Por último, se evaluó el modelo calculando intervalos de confianza simultáneos al 95 % para los coeficientes con el método de Hotelling-Scheffé.

## 2.2. Modelo de riesgos competitivos

El análisis de supervivencia engloba un conjunto de técnicas que permiten analizar la variable *tiempo; desde el inicio del seguimiento hasta que ocurre el evento de interés*. Los modelos de riesgos competitivos son un tipo de análisis de supervivencia apropiado para analizar cómo se comporta un sujeto que puede fallar por diferentes causas distintas al evento de interés, entonces para cada unidad en estudio se observa tanto el tiempo hasta el fallo, como el tipo de evento (fallo) [40, 41].

Los modelos de riesgos competitivos fueron descritos por David Bernoulli en 1760 con el objetivo de describir el riesgo de morir a causa de la viruela o de otras causas. Sin embargo, diversos autores han publicado trabajos en el campo de la estadística y las matemáticas para mejorar esta técnica, entre ellos destaca el trabajo de David y

Moeschberger en 1978 [42]. Actualmente, la aplicación de los modelos de riesgos competitivos es diversa; por ejemplo su aplicación en medicina para el estudio del riesgo de mortalidad por diferentes causas, así como en los ensayos clínicos para encontrar el beneficio de nuevos medicamentos para prevenir ciertos eventos cuando hay otros desenlaces que pueden condicionar que el sujeto falle en el seguimiento en el tiempo [41].

La característica distintiva de un ajuste del modelo de riesgos competitivos es que cada individuo está caracterizado por un par  $(T, C)$ , donde  $T > 0$  es el tiempo desde el inicio del seguimiento hasta la ocurrencia del fenómeno de interés y  $C \in 1, \dots, k$  representa la causa del fallo. Por lo tanto, se requiere un modelo conjunto para  $T$  y  $C$  [41]. Es importante mencionar que el término *fallo* se utiliza para indicar cualquier evento de interés dependiendo el estudio. En cambio, el término *tiempo* no significa necesariamente el tiempo cronológico, sino que puede ser cualquier medida no decreciente en relación con el tiempo cronológico, por ejemplo; número de ciclos, número de kilómetros recorridos, etc. [43].

En ocasiones, se produce una situación de riesgos competitivos cuando la aparición de un evento cualquiera modifica la capacidad de observar el evento de interés. Cuando no se considera la situación previa y se censuran los eventos distintos al de interés, el análisis se reduce a un escenario de tiempo hasta el evento *habitual* y se corre el riesgo de sobrestimar la probabilidad del evento de interés. Por lo tanto, cuando el o los eventos de riesgos competitivos se censuran y se mezclan con las observaciones censuradas verdaderas (tiempo incompleto dado que no se ha producido el evento de interés u otros evento de riesgos competitivos [44]), debe cumplirse que el evento de riesgos competitivos es independiente del evento de interés. En ese caso, pueden interpretarse como el efecto de covariables cuando no existen eventos de riesgos competitivos. Sin embargo, este supuesto muy pocas veces es válido, además de que no es posible verificarlo ni ponerlo a prueba. En resumen, cuando los riesgos competitivos se consideran como observaciones censuradas, la estimación de la probabilidad del

evento de interés será incorrecta y la interpretación del efecto de las covariables no estará clara debido a la falta de conocimiento de la independencia entre el evento de interés y el evento de riesgos competitivos. En cambio, al codificar los riesgos competitivos como un evento distinto al de interés y distinto de las observaciones censuradas verdaderas, la probabilidad se estimará de manera correcta y el modelo tendrá una interpretación más sencilla [40].

### 2.2.1. Especificación del modelo de riesgos competitivos

Se define para cada individuo  $(T, C)$  donde  $T$  es una variable aleatoria continua y positiva, mientras que  $C$  toma valores del conjunto finito  $1, \dots, K$ . Entonces, se asume que cada sujeto falla debido a una y sólo una causa. Por lo tanto, la distribución conjunta de  $(T, C)$  se especifica con cualquiera de las funciones de riesgo causa-específica  $\lambda_j(t)$  o través de las funciones de incidencia acumulada,  $F_j(t)$  [41, 43, 45].

La función de riesgo de causa específica, en el tiempo  $t$  para la  $j$  –ésima causa en presencia de todas las causas actuando simultáneamente, como:

$$\lambda_j(t) = \lim_{\Delta t \rightarrow 0} \left( \frac{Pr(T < t + \Delta t, C = j | T \geq t)}{\Delta t} \right) \quad j = 1, \dots, K \quad (2.15)$$

Entonces,  $\lambda_j(t) = \lim_{\Delta t \rightarrow 0}$  es igual a  $\frac{1}{\Delta t}$  por la probabilidad de morir por la causa  $j$  en  $(t, t + \Delta t)$  y actuando todos los riesgos en  $(t, t + \Delta t)$  dado que sobrevivió a todas las causas. hasta  $t$ . Esto representa la tasa para el  $j$  –ésimo fallo en  $t$  y en presencia de las otras causas, condicionado a que no ha ocurrido ninguno de los eventos de interés hasta ese tiempo. Dicha función también es conocida como función de sub-riesgo [46].

También, se pueden ver a los riesgos competitivos como un caso especial de un modelo multi-estado [41]. El modelo multi-estado tiene una categoría transitoria *vivo* y  $k$  estados denominados *fallo debido a la causa  $j$*  con  $j \in \{1, \dots, k\}$ . Para este modelo, el objetivo es ajustar las probabilidades de la transición de *vivo* a cualquier

evento de interés usando la función  $P_{hj}(s, t)$ , probabilidad de estar en el evento  $j$  en el momento  $t$ , dado que en el momento  $s$  el proceso estaba en el estado  $h$ . Por lo tanto,  $P_{0,t} = P(T \geq t, C = j)$ , las funciones de incidencia acumulada. En cambio, las funciones de intensidad de transición son los riesgos de causa específica [43].

### 2.2.2. Funciones de riesgo causa-específica $\lambda_j(t)$ ; modelo de riesgos proporcionales de Cox

El modelo clásico de riesgos competitivos se basa en un modelo de riesgos proporcionales de Cox (PH) [47]. Entonces, para cada causa-específica la función de riesgo causa específica  $j$ :

$$\lambda_j(t|\mathbf{Z}) = \lambda_{0j} e^{\beta'_j \mathbf{Z}} \quad j = 1, \dots, K \quad (2.16)$$

$\mathbf{Z}$  es un vector de  $p$  covariables y  $\beta'_j$  es un vector de  $p$  coeficientes de regresión para causa específica  $j$ . En este caso, cada evento (fallo) se analiza individualmente, se consideran observaciones censuradas a los individuos que fallan por otra causa distinta al evento de interés o no presentan ningún evento. Entonces, se considera que las covariables actúan multiplicativamente con base a una función de riesgo desconocida  $\lambda_{0j}$ . En el análisis de riesgos proporcionales clásicos (PH), la validez del modelo no depende de la verdadera función de riesgo, siempre y cuando la forma multiplicativa del modelo sea correcta [41, 47].

Utilizando el método de verosimilitudes parciales es posible estimar los parámetros de regresión  $\beta_j$ . Si partimos de que una muestra aleatoria censurada  $(y_i, \delta_i, \delta_i c_i)$  con  $i = 1, \dots, n$ , tiene  $N$  tiempos distintos de fallo observados  $t_1 < \dots < t_N$  y  $n - N$  tiempos de censura. La probabilidad de que un sujeto tenga el evento  $j$  en el tiempo  $t_i$ , dado que uno de los individuos en riesgo (vivo y sin censura) falla en el tiempo  $t$  debido a la causa  $j$ :

$$\frac{e^{\beta'_j \mathbf{Z}_i}}{\sum_{\ell=1}^n \Upsilon_{\ell}(t_i) e^{\beta'_j \mathbf{Z}'_{\ell}}} \quad (2.17)$$

Dado que  $\Upsilon_\ell(t) = I(t_\ell \geq t)$ , entonces la función parcial de verosimilitud únicamente está definida en los tiempo de fallo de la siguiente forma:

$$L(\beta_1, \dots, \beta_k) = \prod_{i=1}^n \prod_{j=1}^k \left( \frac{e^{\beta_j' \mathbf{Z}_i}}{\sum_{\ell=1}^k \Upsilon_\ell(t_i) e^{\beta_\ell' \mathbf{Z}_i}} \right)^{\delta_{ij}} = \prod_{j=1}^k L_j(\beta_j) \quad (2.18)$$

donde  $\delta_{ij} = I(C_i = j)$ . Entonces, el grupo de riesgo puede disminuir por la ocurrencia de un evento por cualquier otra causa.

Al maximizar cada factor en la ecuación anterior nos permite encontrar un estimador  $\hat{\beta}_j$  consistente y asintóticamente normal bajo ciertas condiciones. En cambio, el vector score, la matriz de información y el estadístico de razón de verosimilitudes basados en  $L(\hat{\beta}_j)$  se comportan como si se deducieran de una verosimilitud ordinaria [41, 43, 47].

Debido a que el propósito del modelo de regresión de Cox es evaluar simultáneamente el efecto de varios factores (covariables) sobre la supervivencia. Es decir, permite examinar cómo los factores específicos influyen en la tasa de ocurrencia de un evento particular en un momento determinado. A esta tasa se le conoce comúnmente como la tasa de riesgo (Hazard rate) [48, 49]. Podemos reescribir la *ecuación 2.16* de la siguiente forma:

$$H(t) = H_0(t) e^{(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \quad (2.19)$$

Donde,  $t$  representa el tiempo de supervivencia y  $h(t)$  es a función de riesgo determinada por las covariables  $(x_1, x_2, \dots, x_p)$ . En cambio, los coeficientes  $(\beta_1, \beta_2, \dots, \beta_p)$  miden el impacto de las covariables. Por último, el término  $H_0(t)$  representa el riesgo basal, es decir, corresponde al valor del riesgo si todas las covariables son iguales a 0 y, entonces, la expresión  $e^{(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} = 1$ . Además, al calcular el exponente de los coeficientes del modelo es posible calcular el *Hazard ratio* (HR) de un factor de riesgo (covariable) o predictor dado en el modelo.

El HR es el riesgo de fallo, es decir, el riesgo o probabilidad de sufrir el evento de interés, dado que el participante ha sobrevivido hasta un tiempo determinado.

Aunque, una probabilidad debe estar en el rango de 0 a 1. El HR representa el número esperado de eventos por unidad de tiempo, entonces el HR en un grupo puede exceder la unidad (1) [48, 49]. Considerando específicamente que el HR es la relación entre el número total de eventos observados y esperados en dos grupos independientes, podemos expresar el HR en un modelo de regresión de Cox como:

$$\frac{H(t)}{H_0(t)} = e^{(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \quad (2.20)$$

Al interpretar el HR, debemos considerar que si un HR es superior a 1 indica una covariable que está positivamente asociada con la probabilidad del evento y, por lo tanto, negativamente asociada con la duración de la supervivencia. En resumen:

1.  $HR = 1$ : no hay efecto (asociación).
2.  $HR < 1$ : Reducción de la probabilidad de ocurrencia del evento, reducción del riesgo.
3.  $HR > 1$ : Aumento de la probabilidad de ocurrencia del evento, aumento del riesgo.

Un punto importante a resaltar es que en el modelo de Cox se asume que las curvas del riesgo para los grupos de observaciones (sujetos/eventos) deben ser proporcionales y no cruzarse. Entonces considerando 2 pacientes  $k$  y  $k'$  que difieren en sus valores de  $x$ , la función de riesgos puede expresarse como:

Para el paciente  $k$ :

$$h_k(t) = h_0(t)e^{\sum_{i=1}^n \beta x} \quad (2.21)$$

Para el paciente  $k'$ :

$$h_{k'}(t) = h_0(t)e^{\sum_{i=1}^n \beta x'} \quad (2.22)$$

Por lo tanto, el HR para estos dos pacientes será independiente del tiempo:

$$\frac{h_k(t)}{h_{k'}(t)} = \frac{h_0(t)e^{\sum_{i=1}^n \beta x}}{h_0(t)e^{\sum_{i=1}^n \beta x'}} = \frac{e^{\sum_{i=1}^n \beta x}}{e^{\sum_{i=1}^n \beta x'}} \quad (2.23)$$

Por lo tanto, el modelo regresión de Cox al ser un modelo de riesgos proporcionales: el riesgo del evento en cualquier grupo es un múltiplo constante del riesgo en cualquier otro. Esta suposición implica que, como se mencionó anteriormente, las curvas de riesgo para los grupos deben ser proporcionales y no pueden cruzarse [49]. Por ejemplo, si un individuo tiene un riesgo de muerte en algún momento inicial que es el doble que el de otro individuo, entonces en todos los momentos posteriores el riesgo de muerte sigue siendo el doble.

### 2.2.3. El modelo de Fine y Gray

En 1999 Fine y Gray describieron una nueva función [50], la *función de sub-riesgo*  $y_j(t)$  que deriva de la función de subdistribución:

$$y_j(t|\mathbf{Z}) = \lim_{\Delta t \rightarrow 0} \left( \frac{Pr(T < t + \Delta t, C = j | \mathbf{Z}, \{T \geq t \text{ or } (T < t \text{ and } C \neq j)\})}{\Delta t} \right) \quad (2.24)$$

$$= \frac{f_j(t|\mathbf{Z})}{1 - F_j(t|\mathbf{Z})} \quad j = 1, \dots, k. \quad (2.25)$$

Donde el riesgo obtenido de  $F_j$  es una distribución apropiada. Además, la expresión condicional incluye:

1. El evento no ha ocurrido en el tiempo  $t$ .
2. El evento se ha producido por una causa distinta antes de  $t$

Por lo tanto, el riesgo en el tiempo  $t$  está conformado por dos tipos de sujetos, que corresponde a escenarios distintos. Es importante señalar que a diferencia de los análisis basados en una causa específica, un sujeto que falla por otras causas no sería eliminado del conjunto de riesgos en el momento que su fallo se presenta. Entonces la función de sub-riesgos está dado por:

$$F_j(t|\mathbf{x}) = 1 - \exp \left( - \int_0^t y_i(t|\mathbf{x}) \right) \quad (2.26)$$

Dada la función de su-riesgo, Fine y Gray propusieron adaptarla en un modelo de Cox, entonces:

$$y_i(t|\mathbf{x}) = y_{0j}(t)e^{\beta'_j \mathbf{x}} \quad j = 1, \dots, k. \quad (2.27)$$

Donde las variables actúan linealmente sobre una transformación logarítmica de una complementaria de la función de incidencia acumulada. Además, Fine y Gray demostraron que el método de verosimilitud parcial es válido para la estimación cuando la censura está ausente o siempre es observable. Si hay  $N$  fallos en los tiempos  $t_1 < t_2 < \dots < t_N$ , la verosimilitud parcial se define como:

$$\hat{L}(\beta_j) = \prod_{i=1}^N \left( \frac{e^{\beta'_j \mathbf{Z}_i}}{\sum_{\ell \in R_i} w_{i\ell} e^{\beta'_j \mathbf{Z}'_{\ell}}} \right) \quad (2.28)$$

Donde el conjunto de riesgo para el evento específico  $j$  en el momento  $t_i$  es  $R_i = \{l : t_l \geq t_i \text{ o } (t_l \leq t_i \text{ y } C \neq j)\}$ , entonces lo sujetos que presentan un evento alternativo permanecen en el grupo de riesgo. En cambios, el  $w_{i\ell}$  para cada individuo es  $G(t_i/G(\min(t_l, t_i)))$  y  $G$  es la función de supervivencia para la distribución de la censura. Por lo tanto, un individuo que satisface  $t_l \geq t_i$  tiene un peso de 1 [41, 45, 50].

Por otro lado, en un modelo de riesgo de causa específica el coeficiente de regresión exponenciado denotará la magnitud del cambio relativo en la función de riesgo de causa específica asociado con un cambio de 1 unidad en la covariable. Por lo tanto, el HR denota el cambio relativo en la tasa de ocurrencia del evento primario en sujetos que actualmente están libres de eventos. La tasa de ocurrencia del evento denotará la intensidad con la que ocurren los eventos. Entonces, el HR de causa específica puede interpretarse como un cociente de tasas. Sin embargo, cuando se usa un modelo de riesgo de causa específica en presencia de riesgos competitivos, es incorrecto inferir que una variable determinada está asociada con una incidencia mayor o menor del

evento de interés, ya que la *fórmula 2.16* no se cumple en presencia de riesgos competitivos. Debido a que se debe tener en cuenta el efecto de las covariables en la función de riesgo de causa específica de cada uno de los diferentes tipos de eventos [51, 52].

Entonces, al exponenciar el o los coeficiente de regresión del modelo de riesgos de subdistribución de Fine y Gray (subdistribution Hazard ratio, sHR), este denotará la magnitud del cambio relativo en la función de riesgo de subdistribución asociada a un cambio de 1 unidad en la covariable dada. Es decir, reporta el cambio relativo en la tasa de ocurrencia del evento en aquellos sujetos que están libres de eventos o que han experimentado un evento competitivo. Al considerar esta interpretación, se debe aceptar que aquellos que experimentaron eventos competitivos están exentos (*cura*) del evento principal de interés y que estar en el riesgo establecido después del evento competitivo representa un tiempo "*immortal*". Si bien tales modelos se han adoptado ampliamente en entornos donde los riesgos competitivos no son observables. En riesgos competitivos donde la falla por otras causas es observable, algunos profesionales pueden encontrar esta interpretación difícil de conceptualizar. En consecuencia, puede haber una preferencia por hacer inferencias sobre la magnitud de los efectos de las covariables en la incidencia del resultado [53].

Una alternativa de la interpretación del sHR, es interpretar las covariables como si tuvieran un efecto sobre la incidencia del evento. A pesar de que esta interpretación alternativa puede ser más fácil de entender y comunicar a quienes no están familiarizados con este tipo de análisis, esta interpretación tampoco está exenta de dificultades. La principal dificultad radica en la interpretación numérica del sHR. Por lo tanto, es importante enfatizar que la magnitud del efecto relativo de la covariable en la función de riesgo de subdistribución es diferente de la magnitud del efecto de la covariable en la función de incidencia acumulada. Esto es similar a los datos sin riesgos competitivos, donde la interpretación de la razón de riesgo en términos de probabilidades de supervivencia requiere cuidado. En resumen, si una variable aumenta la función de riesgo de subdistribución, también aumentará la incidencia del evento. Sin embargo,

no se puede inferir que las magnitudes de estos dos efectos sean del mismo tamaño, a pesar de estar en la misma dirección. A pesar de esta limitación, una prueba de significancia estadística de sHR proporciona una prueba del efecto de la covariable en la función de incidencia acumulada. Esto es algo que no es posible mediante el uso del modelo de riesgo de causa específica. Ni la dirección del HR de causa específica ni su significancia estadística proporcionan información sobre la dirección o la significancia de la asociación de la covariable con la función de incidencia acumulada [53, 54].

#### 2.2.4. Implementación del modelo de Fine y Gray

Utilizando el modelo de Fine y Gray, un modelo basado en el modelo de riesgos proporcionales de Cox, se realizó un análisis de riesgos competitivos bajo cada causa de ocurrencia de un evento de interés, por lo tanto se consideró como una función de riesgo en presencia de covariables. Debido a que el número de fallos por causas distintas a la causa de interés puede reducir el número real de fallos por la causa de interés e influir en la estimación de la probabilidad de falla de la causa de interés. Se ajustaron modelos de Fine y Gray, un modelo de regresión de supervivencia utilizando la función de incidencia acumulada y las funciones de riesgo de sub-distribución. Entonces, los parámetros del modelo se estimaron incorporando pesos en la función de verosimilitud parcial. Por ende, para una covariable  $x_i$ , la razón de riesgo de sub-distribución para la causa  $j$  ( $j = 1, \dots, k$ ) está dada por  $exp(\beta_{ij})$  manteniendo todas las demás covariables en un nivel fijo, donde  $\beta_{ij}$  es el coeficiente de la regresión [55].

Para ajustar el modelo de Fine y Gray, se consideraron las siguientes variables:

- **Tiempo de seguimiento:** Se calculó de la siguiente manera: fecha de la aplicación de la encuesta menos la fecha de la presencia del evento (mortalidad por cualquier causa) o la ausencia del mismo (censura), lo que ocurriera primero.
- **Mortalidad cardiovascular:** Codificado como *mortalidad cardiovascular*, *mortalidad por otra causa* o *censurado*.
- **Covariables:** se incluyeron en el modelo las siguientes variables; *HOMA2IR*, *re-*

manentes de colesterol, sexo, edad, IMC, etnicidad, tabaquismo, colesterol LDL y número de comorbilidades. Para el caso de remanentes de colesterol, se incluyó la estimación dada por la fórmula de Sampson y Martin (Rem-S y Rem-M). Por lo tanto, se ajustaron 2 modelos en cada submuestra (sujetos sin diabetes y sujetos con diabetes).

El ajuste del modelo se realizó utilizando las siguientes líneas de código en R:

```
fgdata_cv <- finegray(Surv(tiempo, evento) ~ .,
                     data=base, na.action=na.pass)
m1 <- coxph(Surv(fgstart, fgstop, fgstatus)~HOMA2IR+Remanentes+
            Sexo+edad+etnicidad+IMC+comorbilidades,
            weight=fgwt,
            data=fgdata_cv)
```

### 2.2.5. Evaluación del modelo de Fine y Gray

Para evaluar si el modelo de Fine y Gray, un modelo basado en el modelo de riesgos proporcionales de Cox, describe adecuadamente los datos se evaluó;

1. Evaluación de la suposición de riesgos proporcionales: para comprobar este supuesto se evalúan los residuos de Schoenfeld. Estos residuos deben ser independientes del tiempo, una representación gráfica que muestre un patrón no aleatorio contra el tiempo es evidencia de violación del supuesto. Al usar pruebas estadísticas para evaluar este supuesto, una relación no significativa respalda la suposición de riesgos proporcionales.
2. Examinación de datos influyentes: evaluación de la devianza residual a través de la transformación de los residuales de Martingala. Estos residuos deben distribuirse simétricamente alrededor del cero con una desviación estándar de 1. Valores positivos indican sujetos que tuvieron el fallo pronto en comparación con los tiempos de supervivencia esperados. En cambio, valores negativos describen a sujetos que vivieron demasiado. Por lo tanto, valores muy grandes o pequeños son valores atípicos que el modelo predice de manera deficiente.

3. Evaluación de la no linealidad en la relación entre el logaritmo del riesgo y las covariables: evaluación de los residuales de Martingala para evaluar la no linealidad. La no linealidad no es un problema para las variables categóricas, por lo que sólo se evalúan los gráficos de residuos Martingala y residuos parciales frente a las variables continuas. La relación debe ser lineal para satisfacer los supuestos del modelo de riesgos proporcionales.

### 2.3. Modelo de Mediación Causal (AMC)

Muchas veces en investigación clínica, la relación entre variables es compleja. Usualmente, se utiliza el modelo de RLM para explorar las relaciones entre las variables explicativas y la variable dependiente. Sin embargo, en muchas ocasiones el modelo de RLM no logra dilucidar los mecanismos subyacentes a la asociación entre la variable explicativa y la variable dependiente. Además, en ciertas ocasiones los investigadores están interesados describir los procesos fisiopatológicos mediadores subyacentes. Partiendo del conocimiento previo de que una variable explicativa ejerce su efecto sobre el resultado a través de vías directas e indirectas. En la vía indirecta, generalmente puede establecerse que hay un mediador que transmite el efecto causal [56, 57, 58].

Considerando un modelo simple con las variables  $X_i$  (independiente) e  $Y_i$  (dependiente). La mediación en su forma más simple se logra al agregar un mediador  $M$  entre  $X_i$  e  $Y_i$ , *Figura 2.2*. El análisis de mediación basado en variables con resultado continuo, que se basa en tres ecuaciones lineales:

$$Y_i = \beta_1 + cX_i + e_{1i} \quad (1) \quad (2.29)$$

$$Y_i = \beta_2 + c'X_i + bM_i + e_{2i} \quad (2) \quad (2.30)$$

$$M_i = \beta_3 + aX_i + e_{3i} \quad (3) \quad (2.31)$$

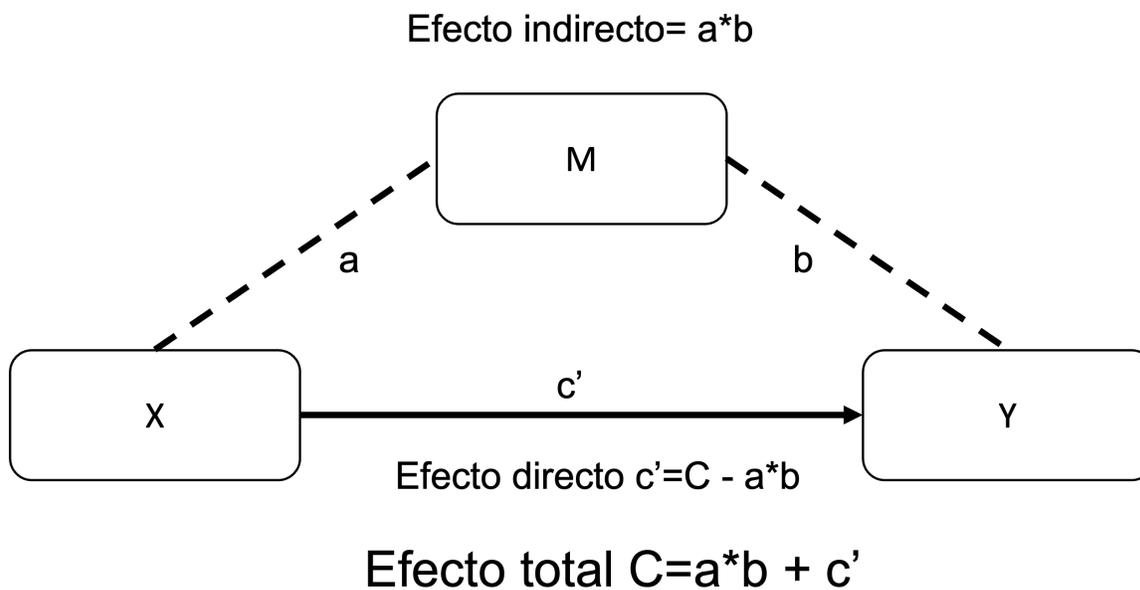


Figura 2.2: Esquema general de un modelo de mediación causal

Donde  $c$  es el coeficiente para el efecto de  $X_i$  en  $Y_i$  que representa el efecto causal total.  $c'$  es el coeficiente para el efecto que relaciona a  $X_i$  y  $Y_i$  ajustado por el mediador ( $M_i$ ) que describe el efecto directo. En cambio,  $b$  es el coeficiente para el efecto de  $M_i$  sobre  $Y_i$  ajustado por  $X_i$  y  $a$  es el coeficiente relativo al efecto  $X_i$  sobre  $M_i$ . Además,  $e_1, e_2, e_3$  son los residuales que no están correlacionados y son independientes entre sí. Bajo esto, el efecto de mediación causal (efecto indirecto) está representado por el coeficiente del producto de  $a*b$ . Por lo tanto la ecuación 3 puede sustituirse en la ecuación 2 para eliminar el término  $M_i$ :

$$Y_i = \beta_2 + b\beta_3 + (c' + ab)X_i + e_{2i} + be_{3i} \quad (2.32)$$

Los parámetros relacionados con el efecto directo ( $c'$ ) e indirecto ( $ab$ ) de  $X_i$  sobre  $Y_i$  son diferentes a los de su efecto total. Es decir, probar la hipótesis nula  $c = 0$  es innecesario ya que el efecto de mediación puede ser distinto de cero incluso cuando el efecto causal total es cero (es decir, los efectos directos e indirectos pueden ser opuestos), lo que refleja la cancelación del efecto de diferentes caminos [56].

### 2.3.1. Análisis de mediación causal en análisis de supervivencia

Interesantemente en los datos de supervivencia dentro del contexto del análisis de mediación causalmente ordenados existen múltiples formas o escalas mediante las cuales es posible dilucidar un efecto total comparando los niveles de exposición en efectos directos e indirectos [59]. Además, la presencia de censura modifica las propiedades de los estimadores utilizados en los casos evaluados con datos no censurados.

Si se considera que  $A$  denota la exposición de interés,  $T$  es un resultado tiempo-evento,  $M$  el mediador y  $C$  es un conjunto de covariables. Entonces,  $Ta$  denota el tiempo del evento contrafáctico si  $A$  se fija en  $a$ ; además,  $Tam$  denota el tiempo del evento contrafáctico si  $A$  se hubiera fijado en  $a$  y  $M$  se fija en  $m$ . En cambio, para una variable arbitraria de tiempo hasta el evento  $V$ ,  $SV(t)$  denota la función de supervivencia en el tiempo  $t$ , es decir,  $SV(t) = P(V > t)$ ; la función de supervivencia condicionada a las covariables  $C = c$ , también se puede definir como  $SV(t|c) = P(V > t|c)$ . Por lo tanto,  $\lambda V(t)$  y  $\lambda V(t|c)$  para el riesgo o riesgo condicional en el tiempo  $t$ , que es la tasa instantánea del evento condicional a  $V \geq t$  [59, 60].

Si se considera a las funciones de supervivencia, es posible comparar las funciones  $STa(t)$  y  $STa^*(t)$ :

$$STa(t) - STa^*(t) = \left( S_{T_{aMa}} - S_{T_{aMa^*}} \right) + \left( S_{T_{aMa^*}} - S_{T_{a^*Ma^*}} \right) \quad (2.33)$$

La primera expresión dentro del paréntesis representa el efecto indirecto en la escala de función de supervivencia, en cambio la segunda expresión indica el efecto directo. Además, es posible representar la diferencia general en los riesgos (hazards) como la suma de los efectos directo e indirecto en la escala de riesgos (Hazard), tal como se muestra:

$$\lambda_{Ta}(t) - \lambda_{Ta^*}(t) = \left( \lambda_{T_{aMa}} - \lambda_{T_{aMa^*}} \right) + \left( \lambda_{T_{aMa^*}} - \lambda_{T_{a^*Ma^*}} \right) \quad (2.34)$$

Lange y Hansen consideraron ambas expresiones además de la descomposición del efecto de riesgo acumulativo [60]. Pero, también es posible considerar otras representaciones de los efectos. Es así que, es posible considerar estas representaciones en términos de tiempos medios de supervivencia, tal como se presenta:

$$E(T_a) - E(T_{a*}) = \left( E(T_{aMa}) - E(T_{aMa*}) \right) + \left( E(T_{aMa*}) - E(T_{a*Ma*}) \right) \quad (2.35)$$

Además, es posible usar la diferencia en función de supervivencia logarítmica, o riesgos (Hazard) logarítmicos, etc. Por ejemplo, con riesgo (Hazard) logarítmico se tiene:

$$\log\{\lambda_{T_a}(t)\} - \log\{\lambda_{T_{a*}}(t)\} = \left( \log\{\lambda_{T_{aMa}}\} - \log\{v\lambda_{T_{aMa*}}\} \right) + \left( \log\{\lambda_{T_{aMa*}}\} - \log\{\lambda_{T_{a*Ma*}}\} \right) \quad (2.36)$$

Que al exponenciar dicha expresión se obtiene:

$$\lambda_{T_a}(t)/\lambda_{T_{a*}}(t) = \left( \lambda_{T_{aMa}}/\lambda_{T_{aMa*}} \right) * \left( \lambda_{T_{aMa*}}/\lambda_{T_{a*Ma*}} \right) \quad (2.37)$$

Por lo tanto, la relación de riesgos es el producto de las relaciones naturales del efecto directo e indirecto. Además, las medidas anteriores se pueden considerar condicionales a los estratos de covariables  $C = c$ . Considerando las expresiones anteriores es posible calcular la *proporción mediada* tomando la relación entre el efecto indirecto y la suma del efecto indirecto y directo, es decir, el efecto total [59]. Otra forma de interpretar el modelo en términos de supervivencia es que el efecto indirecto es el número de muertes que se pueden atribuir a la mediación a través del mediador, mientras que el efecto directo es el número de muertes que se pueden atribuir a una vía directa (o a otros mediadores no incluidos en el análisis). Entonces, el efecto total será el número de muertes causadas tanto por el efecto directo como el indirecto [60].

Sin embargo, para poder obtener una conclusión causal de un análisis se deben

de cumplir una serie de suposiciones sobre la ausencia de factores de confusión no medidos, también denominados suposiciones de intercambiabilidad. Por lo tanto, se debe suponer que no hay factores de confusión no medidos para la relación exposición-evento, el mediador-evento y exposición-mediador. Además, se debe satisfacer la condición de identificabilidad de Pearl que indica que el efecto de la exposición tiene su efecto a través de 2 vías causales distintas y no entrelazadas. Dicha suposición se violaría si una variable se ve afectada por la exposición y afecta tanto al mediador como al evento de interés. Por último, se incluye el supuesto de consistencia que implica que el resultado no se altera si se establece la exposición y el mediador en los valores que tomarían naturalmente [60, 61, 62].

### 2.3.2. Análisis de mediación causal en modelos de riesgos proporcionales

Tein y Mackinnon por primera vez aplicaron el modelo de mediación al análisis de supervivencia en el ámbito de ciencias sociales. Existen dos métodos para llevar a cabo el análisis de mediación: el *método de la diferencia*, común en epidemiología, considera un modelo que incluye tanto al mediador como la ausencia del mismo para estimar la diferencia en los coeficientes de la exposición como medida del efecto indirecto o mediado. En cambio, el *método del producto*, común en las ciencias sociales, toma como medida del efecto indirecto el producto del coeficiente de exposición en el modelo para el mediador y el coeficiente para el mediador en el modelo para el resultado. Si el resultado y el mediador son continuos y no hay interacciones en el modelo para el resultado, entonces los dos métodos coinciden. Sin embargo, con resultados binarios, los dos métodos pueden divergir, pero coincidirán aproximadamente cuando el resultado binario sea raro [59, 63, 64, 65].

Tein y MacKinnon utilizando simulaciones, demostraron que el método de la diferencia y el método del producto dan resultados diferentes para el modelo de riesgos proporcionales. Por lo tanto, para el modelo de mediación en riesgos proporcionales utilizaron:

$$\lambda_T(t|a, m, c) = \lambda_T(t|0, 0, 0)e^{\gamma_1 a + \gamma_2 m + \gamma_4 c} \quad (2.38)$$

Tal como ha sido demostrado por VanderWeele [59], cuando el evento de interés es raro, los efectos indirectos y directos están dados por:

$$\log\{\lambda_{T_{a_{Ma}}}(t|c)\} - \log\{\lambda_{T_{a_{Ma^*}}}(t|c)\} = (\gamma_2\beta_1 + \gamma_3\beta_1 a)(a - a^*) \quad (2.39)$$

$$\log\{\lambda_{T_{a_{Ma}}}(t|c)\} - \log\{\lambda_{T_{a_{Ma^*}}}(t|c)\} = \{\gamma_1 + \gamma_3(\beta_0 + \beta_1 a^* + \beta_2' c + \gamma_2 \sigma^2)\}(a - a^*) + 0.5\gamma_3^2 \sigma^2 (a^2 - a^{*2}) \quad (2.40)$$

Donde  $\sigma^2$  es la varianza del término de error en el modelo de regresión para el mediador. Dichas expresiones son equivalentes a las obtenidas para resultados dicotómicos, pero sólo cuando el evento de interés es raro. Los coeficientes de los efectos directo e indirecto se obtienen exponenciando el lado derecho de las expresiones [59].

Considerando el escenario donde el evento de interés no es raro, Tein y MacKinnon han demostrado que ambos métodos (diferencia y producto) pueden divergir e incluso ser de signos opuestos. Además, Lange y Hansen concluyeron que en el modelo de riesgos proporcionales, los efectos directo e indirecto no tiene una expresión analítica simple. Sin embargo, VanderWeele describió que cuando el evento es común, el método del producto proporciona una prueba válida de si hay algún efecto mediador siempre y cuando los modelos estén correctamente especificados. Por lo tanto, con el modelo de riesgos proporcionales y un evento común, el método del producto puede ser útil para probar la hipótesis de cualquier efecto mediador. Sin embargo, el método del producto ni el de la diferencia deberían utilizarse en general como medida de un efecto indirecto [59, 60, 66]

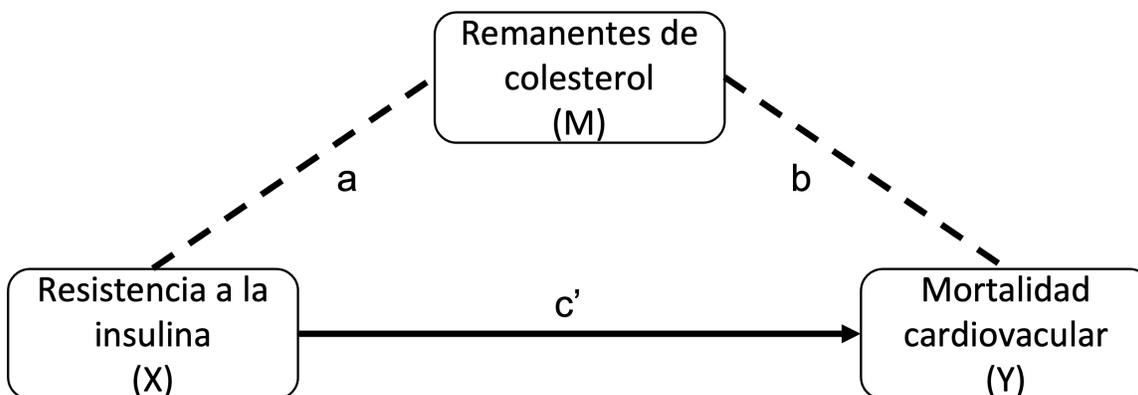


Figura 2.3: Esquema del modelo de mediación causal para evaluar el efecto mediador de los remanentes de colesterol sobre la relación entre resistencia a la insulina y mortalidad cardiovascular.

### 2.3.3. Implementación del modelo de mediación causal: mediadores causalmente ordenados

Considerando el modelo de mediación causal que se presenta en la *figura 2.3*, los 6 supuestos del modelo se mantendrían. Específicamente, se asegura que no hay confusión para las relaciones entre  $X$  y  $Y$ , efecto directo; entre  $X$  y  $M$ , y  $M$  e  $Y$ , efecto indirecto. Además, se asume que no hay otro camino alternativo desde  $X$  hasta  $Y$  y a través de  $M$  hasta  $Y$ . También, se asume que no existe otro mediador desconocido que condicione la relación  $X \dashrightarrow Y$  ni a través de  $X \dashrightarrow M \dashrightarrow Y$ .

Para ello, se ajustaron los siguientes modelos:

- Modelo de RLM:** Se ajustó un modelo de RLM para evaluar la relación del Mediador (Remanentes de colesterol) y la resistencia a la insulina (HOMA2IR), ajustado por las siguientes covariables (edad, sexo, IMC, etnicidad, tabaquismo, colesterol LDL y número de comorbilidades), tal y como se describió en las secciones previas.
- Modelo de riesgos competitivos de Fine y Gray:** se ajustó el modelo de riesgos competitivos para evaluar el riesgo de mortalidad cardiovascular atribuible a la resistencia a la insulina y los remanentes de colesterol, ajustado por edad, sexo, IMC, etnicidad, tabaquismo, colesterol LDL y número de comorbi-

lidades, descrito en la sección previa.

Los análisis de mediación se realizaron utilizando las siguientes líneas de código, para permitir la inferencia de los parámetros estimados en los modelos de mediación, obtuvimos intervalos de confianza del 95% utilizando percentiles de bootstrap no paramétrico acelerado con corrección de sesgo ( $B = 1000$ ). Los análisis de mediación causal para la mortalidad se probaron utilizando un análisis de mediación basado en un modelo ordenado causalmente utilizando coeficientes beta extraídos de los modelos de regresión de riesgos competitivos de Fine y Gray. El código de R utilizado para el ajuste del modelo de mediación causal se presenta a continuación [67, 68]:

```
## method=="Aalen" -> output Hazard Difference
## method=="Cox"   -> output Hazard Ratio
mediation_ci1 <- function(lambda.s, lambda.g, covar11, covar12,
                          covar22, alpha.s, var_alpha,
                          G=10^4, method){

  require(mvtnorm)
  Omega <- matrix(c(covar11,covar12,covar12,covar22),nrow=2)
  IE <- rep(0,G); DE <- rep(0,G); TE <- rep(0,G);
  Q <- rep(0,G)

  set.seed(137)
  lambda <- rmvnorm(G, mean = c(lambda.s, lambda.g),
                   sigma = Omega)
  alpha <- rnorm(G, mean=alpha.s, sd=sqrt(var_alpha))
  DE <- lambda[,1]
  IE <- lambda[,2] * alpha
  TE <- IE + DE

  DE.obs <- lambda.s
  IE.obs <- lambda.g * alpha.s
  TE.obs <- DE.obs+IE.obs
  pval.DE<-2*min(mean((DE-mean(DE))>DE.obs),
```

```

        mean((DE-mean(DE))<DE.obs))
pval.IE<-2*min(mean((IE-mean(IE))>IE.obs),
              mean((IE-mean(IE))<IE.obs))
pval.TE<-2*min(mean((TE-mean(TE))>TE.obs),
              mean((TE-mean(TE))<TE.obs))

if (method=="Cox") {DE=exp(DE); IE=exp(IE); TE=exp(TE)}
print("DE:")
print(ifelse(method=="Aalen", DE.obs, exp(DE.obs)))
print(quantile(DE, c(0.025, 0.975)))
print(paste("pval_DE=", pval.DE))
print("IE:")
print(ifelse(method=="Aalen", IE.obs, exp(IE.obs)))
print(quantile(IE, c(0.025, 0.975)))
print(paste("pval_IE=", pval.IE))
print("TE:")
print(ifelse(method=="Aalen", TE.obs, exp(TE.obs)))
print(quantile(TE, c(0.025, 0.975)))
print(paste("pval_TE=", pval.TE))
}
mediation_ci2 <- function(lambda.s, lambda.m, lambda.g,
                          Sigma.lambda,
                          alpha.s, alpha.m, Sigma.alpha,
                          delta.s, Sigma.delta,
                          G=10^4, method){
  require(mvtnorm)
  SY <- rep(0,G); SGY <- rep(0,G); SMY <- rep(0,G);
  TE <- rep(0,G)
  set.seed(137)
  lambda <- rmvnorm(G, mean = c(lambda.s, lambda.m, lambda.g),
                   sigma = Sigma.lambda)
  alpha <- rmvnorm(G, mean = c(alpha.s, alpha.m),

```

```

        sigma = Sigma.alpha)

delta <- rnorm(G, mean=delta.s, sd=sqrt(Sigma.delta))
SY <- lambda[,1]
SGY <- lambda[,3] * alpha[,1]
SMY <- (lambda[,2] + lambda[,3]*alpha[,2])*delta
TE <- SY+SGY+SMY

SY.obs <- lambda.s
SGY.obs <- lambda.g * alpha.s
SMY.obs <- (lambda.m + lambda.g*alpha.m)*delta.s
TE.obs <- SY.obs+SGY.obs+SMY.obs
pval.SY<-2*min(mean((SY-mean(SY))>SY.obs),
               mean((SY-mean(SY))<SY.obs))
pval.SGY<-2*min(mean((SGY-mean(SGY))>SGY.obs),
                mean((SGY-mean(SGY))<SGY.obs))
pval.SMY<-2*min(mean((SMY-mean(SMY))>SMY.obs),
                mean((SMY-mean(SMY))<SMY.obs))
pval.TE<-2*min(mean((TE-mean(TE))>TE.obs),
                mean((TE-mean(TE))<TE.obs))

if (method=="Cox") {SY=exp(SY); SGY=exp(SGY); SMY=exp(SMY);
  TE=exp(TE)}

print("SY:")
print(ifelse(method=="Aalen", SY.obs, exp(SY.obs)))
print(quantile(SY, c(0.025, 0.975)))
print(paste("pval_SY=", pval.SY))
print("SGY:")
print(ifelse(method=="Aalen", SGY.obs, exp(SGY.obs)))
print(quantile(SGY, c(0.025, 0.975)))
print(paste("pval_SGY=", pval.SGY))
print("SMY:")
print(ifelse(method=="Aalen", SMY.obs, exp(SMY.obs)))
print(quantile(SMY, c(0.025, 0.975)))

```

```
print(paste("pval_SMY=", pval.SMY))
print("TE:")
print(iffelse(method=="Aalen", TE.obs, exp(TE.obs)))
print(quantile(TE, c(0.025, 0.975)))
print(paste("pval_TE=", pval.TE))
}
#### MODELO ####
set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
} else if(method=="Cox"){
  lambdas<-m1_s$coef
  lambdas.var<-m1_s$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2],
              lambdas.var[2,1], lambdas.var[1,1],
              m_S$coef[2], vcov(m_S)[2,2], G=10^6,
              method=method)
```

# Capítulo 3

## Resultados

### 1. Análisis descriptivo: características clínicas, bioquímicas, antropométricas y mortalidad.

Los datos analizados se obtuvieron de 12, 635 sujetos tal y como se describió en la *figura 2.1*. Se incluyeron en el análisis una mayor proporción de hombres (52.3%) con una mediana de edad de 45 (IQR 31-65) años cumplidos y con IMC de 26.2 (IQR 23.1-29.9)  $Kg/m^2$ . En la *Tabla 1* se muestran las características basales de los sujetos incluidos en el análisis.

Al comparar a los sujetos que murieron por cualquier causa contra aquellos que aún continúan en seguimiento, se observó un mayor número de comorbilidades en el grupo que presentó el evento (muerte por cualquier causa),  $p < 0.001$ . Además, este grupo tuvo una edad mayor, una mayor proporción fueron hombres, así como un índice cintura-estatura (ICE) mayor, indicativo de un mayor grado de obesidad central,  $p < 0.001$ . Así mismo, se observaron mayores concentraciones de remanentes de colesterol y demás parámetros del perfil de lípidos. Aunado a un mayor HOMA2IR, indicativo de un mayor grado de resistencia a la insulina,  $p < 0.001$ .

Características basales				
Variable	Total (n=12,635)	Vivos (n=7,590)	Mortalidad por cualquier causa (n=5,045)	p-valor
Sexo (Mujer)	6027 (47.7)	3377 (56.0)	2650 (44.0)	<0.001
Edad (años)	45 (31-65)	35 (27-45)	69 (57-77)	<0.001
IMC (Kg/m <sup>2</sup> )	26.2 (23.1-29.9)	26.0 (22.9-29.8)	26.4 (23.3-30.0)	<0.001
Seguimiento (meses)	272 (172-301)	294 (274-310)	136 (73-210)	<0.001
Caucásicos (%)	5441 (43.1)	2710 (35.7)	2731 (54.1)	<0.001
México- Americanos (%)	3401 (26.9)	2366 (31.2)	1036 (20.5)	<0.001
Afroamericanos (%)	3222 (26.3)	2162 (28.5)	1160 (23.0)	<0.001
Otras etnicidades (%)	470 (3.7)	352 (4.6)	118 (2.3)	<0.001
Tabaquismo (%)	3417 (27.0)	2087 (27.5)	1330 (26.4)	0.160
Diabetes (%)	1032 (8.2)	248 (3.3)	784 (15.5)	<0.001
Hipertensión (%)	3492 (27.8)	1363 (18.1)	2129 (42.4)	<0.001
Insuficiencia cardiaca (%)	481 (3.8)	65 (0.9)	416 (8.6)	<0.001
EVC (%)	376 (3.0)	49 (0.6)	327 (6.5)	<0.001
ICE (%)	0.55 (0.49-0.61)	0.54 (0.48-0.59)	0.58 (0.53-0.64)	<0.001

Características basales				
Variable	Total (n=12,635)	Vivos (n=7,590)	Mortalidad por cualquier causa (n=5,045)	p-valor
Glucosa (mg/dL)	92.0 (85.0-101.0)	90.0 (83.0-97.0)	97.0 (89.0-109.0)	<0.001
HbA1c (%)	5.3 (5.0-5.7)	5.2 (4.9-5.5)	5.6 (5.2-6.1)	<0.001
Péptido C <i>mmol/L</i>	0.65 (0.38-1.01)	0.57 (0.35-0.88)	0.79 (0.49-1.18)	<0.001
Insulina (mUI/mL)	9.1 (6.4-13.7)	8.7 (6.1-12.9)	9.8 (6.8-15.3)	<0.001
HOMA2-IR	1.5 (0.9-2.3)	1.3 (0.8-2.0)	1.8 (1.1-2.8)	<0.001
Colesterol total (mg/dL)	202 (175-233)	195 (170-224)	215 (186-245)	<0.001
Colesterol HDL (mg/dL)	49 (40-59)	50 (41-60)	48 (39-59)	<0.001
Triglicéridos (mg/dL)	114 (80-171)	103 (73-155)	134 (93-193)	<0.001
Colesterol no- HDL(mg/dL)	151 (123-183)	143 (117-173)	164 (135-196)	<0.001
ApoB (mg/dL)*	104 (87-123)	99 (83-118)	111 (95-130)	<0.001
Colesterol LDL (mg/dL)	126.2 (102.2-153.0)	120.9 (98.2-118.0)	135.5 (110.2-162.5)	<0.001
Colesterol LDL-M (mg/dL)	126.1 (102.2-152.6)	120.2 (97.6-145.3)	135.9 (110.6-162.7)	<0.001

Características basales				
Variable	Total (n=12,635)	Vivos (n=7,590)	Mortalidad por cualquier causa (n=5,045)	p-valor
Rem-S (mg/dL)	20.4 (14.5-30.7)	18.5 (13.6-27.7)	23.8 (16.5-34.8)	<0.001
Rem-M (mg/dL)	22.6 (17.3-30.5)	20.6 (16.3-27.9)	25.4 (19.4-33.5)	<0.001

Tabla 1: Características basales de los sujetos incluidos en NHANES III. Los datos se expresan como mediana (rango intercuartil) para el caso de las variables cuantitativas y como frecuencias y porcentajes para las variables cualitativas. Para comparar las características cualitativas de los sujetos se usó la prueba de Chi-Cuadrada, en cambio para comparar las características cuantitativas se utilizó la prueba U de Mann-Whitney. IMC: índice de masa corporal; EVC: enfermedad cerebrovascular; ICE: índice cintura-estatura; Rem-S: estimación realizada con la fórmula de Sampson; Rem-M: estimación realizada con la fórmula de Martín. \* n=7,797.

Tal y como se describió anteriormente, se observó que el grupo que presentó el evento (mortalidad por cualquier causa) tuvieron mayor carga de enfermedades cardiometabólicas. Además, la resistencia a la insulina es un factor importante para el desarrollo de diabetes, por ello es esperado que pacientes con diabetes presenten un HOMA2IR muy alto en comparación con pacientes sin diabetes. Por lo tanto, se optó por segmentar la base para analizar por separado al grupo de sujetos sin diabetes con respecto al grupo de sujetos con diabetes.

Al evaluar la incidencia en la muestra total, la tasa de mortalidad por cualquier causa fue de 10.3 (IC95% : 10.0-10.6) por cada 1000 personas-año, en cambio la mortalidad cardiovascular fue de 3.3 (IC95% : 3.1-3.4) por cada 1000 personas-año. Considerando a las dos submuestras (con y sin diabetes), se realizó el cálculo de

la incidencia de mortalidad por cualquier causa y causa-específica por cada 1000 personas-año, tal y como se muestra en la siguiente tabla.

Mortalidad	Eventos (No diabetes)	Incidencia (No diabetes)	Eventos (Diabetes)	Incidencia (Diabetes)
General	3863 (35.1)	8.7 (8.5-9.0)	1182 (72.8)	25.4 (23.9-26.8)
Cardiovascular	1180 (10.7)	1.8 (1.7-1.9)	421 (25.9)	6.0 (5.5-6.6)
Diabetes	35 (0.3)	0.08 (0.05-0.10)	134 (8.3)	2.9 (2.4-3.4)
Neoplasias	874 (7.9)	1.9 (1.8-2.1)	185 (11.4)	3.9 (3.4-4.5)
EPOC	169 (1.5)	0.38 (0.32-0.44)	29 (1.8)	0.62 (0.39-0.85)
Accidentes	122 (1.1)	0.27 (0.22-0.32)	25 (1.5)	0.53 (0.32-0.74)
Alzheimer	99 (0.9)	0.22 (0.18-0.27)	25 (1.5)	0.53 (0.32-0.74)
Pneumonia	119 (1.1)	0.27 (0.22-0.31)	34 (2.1)	0.73 (0.49-0.98)
Enfermedad renal	46 (0.4)	0.10 (0.07-0.13)	30 (1.8)	0.64 (0.41-0.88)
Otras causas	1192 (10.8)	2.7 (2.5-2.8)	290 (17.9)	6.2 (5.5-6.9)

Tabla 2: Tasas de incidencia por cada 1000 personas-año. Los eventos se presentan como frecuencias y porcentajes, en cambio la incidencia como tasas de incidencia por cada 1000 personas-año con su respectivo intervalo de confianza al 95 %.

Como se observa en la *Tabla 2*, las tasas de mortalidad fueron más altas en sujetos con diabetes. Destacando una mayor tasa de mortalidad por cualquier causa 25.4 (23-9-26.8) por cada 1000 personas-año en comparación con 8.7 (8.5-9.0) por cada 1000 personas-año del grupo sin diabetes. De igual forma, la mortalidad cardiovascular fue mayor en el grupo de sujetos con diabetes; 6.0 (5.5-6.6) vs 1.8 (1.7-1.9).

Posteriormente, se evaluó la distribución de las variables cuantitativas continuas, así como la correlación entre las mismas. En primer lugar, se presenta al grupo de sujetos sin diabetes, *Figura 3.1*

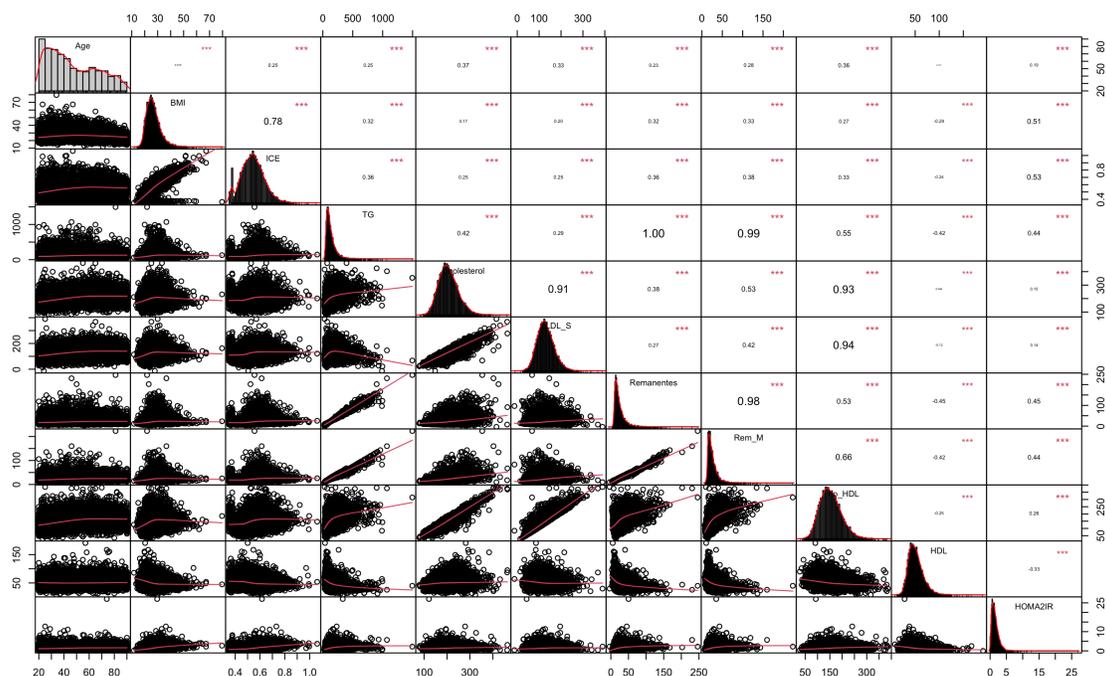


Figura 3.1: Gráficos descriptivos de las variables continuas en sujetos sin diabetes.

Como se muestra en la *Figura 3.1* se observan valores extremos HOMA2IR y en los niveles de remanentes de colesterol calculados con la fórmula de Sampson y Martin, además de que la variable edad no se ajusta a una distribución normal. Considerando que las fórmulas para calcular colesterol LDL, como lo son Martin y Sampson, y que a partir de este se estimaron los remanentes de colesterol, muestran un bajo rendimiento en hipertrigliceridemia severa, se decidió eliminar del análisis a los sujetos con triglicéridos  $\geq 800\text{mg/dL}$  debido a que ambas fórmulas pierden capacidad predictiva a partir de ese nivel de triglicéridos [36]. Además, al observarse valores extremos de HOMA2IR, mayores a 20.0 (percentila 99), se decidió eliminar a los sujetos ya que valores extremos de HOMA2IR pueden indicar que los sujetos tienen una enfermedad asociada a resistencia extrema a la insulina (por ejemplo; lipodistrofias, etc) donde

la relación a estudiar (Remanentes y resistencia a la insulina) no se comporte como en sujetos sin dichas alteraciones metabólicas, además con esto se logró mejorar la distribución de la variable.

También, se decidió evaluar transformaciones para mejorar la simetría de las distribuciones previo al análisis. Considerando que los modelos RLM como los modelos de aprendizaje de máquina tienen mejor desempeño con distribuciones centradas y simétricas [69, 70].

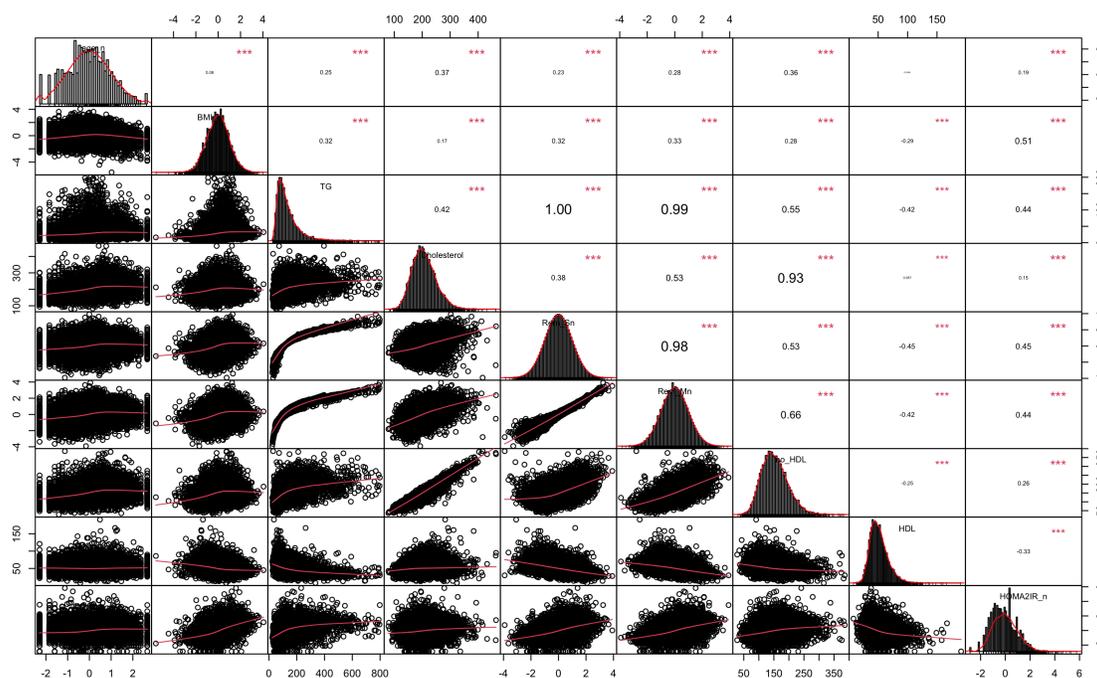


Figura 3.2: Gráficos descriptivos de las variables continuas normalizadas en sujetos sin diabetes.

Utilizando el paquete *BestNormalize* de *R*, se transformaron las variables edad, IMC, remanentes de colesterol y HOMA2IR para disminuir el sesgo y hacer las distribuciones más simétricas. Para el caso de la remanentes de colesterol calculados con la fórmula de Sampson y Martin, y para la edad la mejor transformación fue la normalización de cuantil ordenado (OrderNorm), en cambio para el IMC lo fue la transformación de Box-Cox y para HOMA2IR la raíz cuadrada ( $x + a$ ). La edad fue la variable que no logró una simetría adecuada, pero se optó por trabajar con la transformación de la variable, *Figura 3.2*.

En el grupo de sujetos con diabetes, se observó un compartamiento similar de las variables cuantitativas. Si bien, la edad mostró una distribución un poco más simétrica en comparación con los sujetos sin diabetes. Se observaron niveles extremos de triglicéridos ( $\geq 2500mg/dL$ ) lo que podría afectar la estimación de los remanentes de colesterol, así como valores extremos de HOMA2IR ( $\geq 100$ ), *Figura 3.3*.

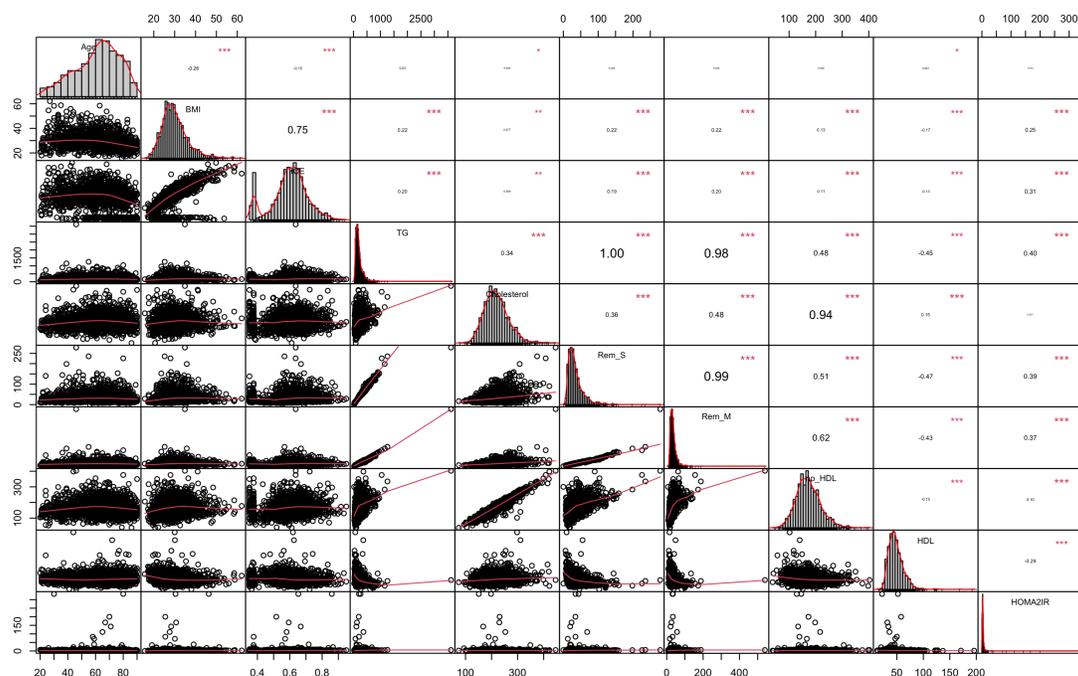


Figura 3.3: Gráficos descriptivos de las variables continuas en sujetos con diabetes.

Por lo tanto, se decidió eliminar del análisis a los sujetos con triglicéridos  $\geq 800mg/dL$  debido al aumento del error de estimación de las fórmulas a partir de ese nivel de triglicéridos, así como aquellos con un HOMA2IR  $\geq 20$  que correspondían al percentil 99. Finalmente, se procedió a transformar a las variables de interés que mostraron mayor asimetría. Para las variables edad, HOMA2IR y remanentes de colesterol calculados con Sampson, la transformación que mejor ajustó los datos fue la normalización de cuantil ordenado (OrderNorm). Mientras que para Remanentes calculados con la fórmula de Martin fue la transformación estandarizada de Yoe-Johnson. Con las transformación se logró mejorar la simetría de las distribuciones de

las variables que se incluirán en los modelos, *Figura 3.4*.

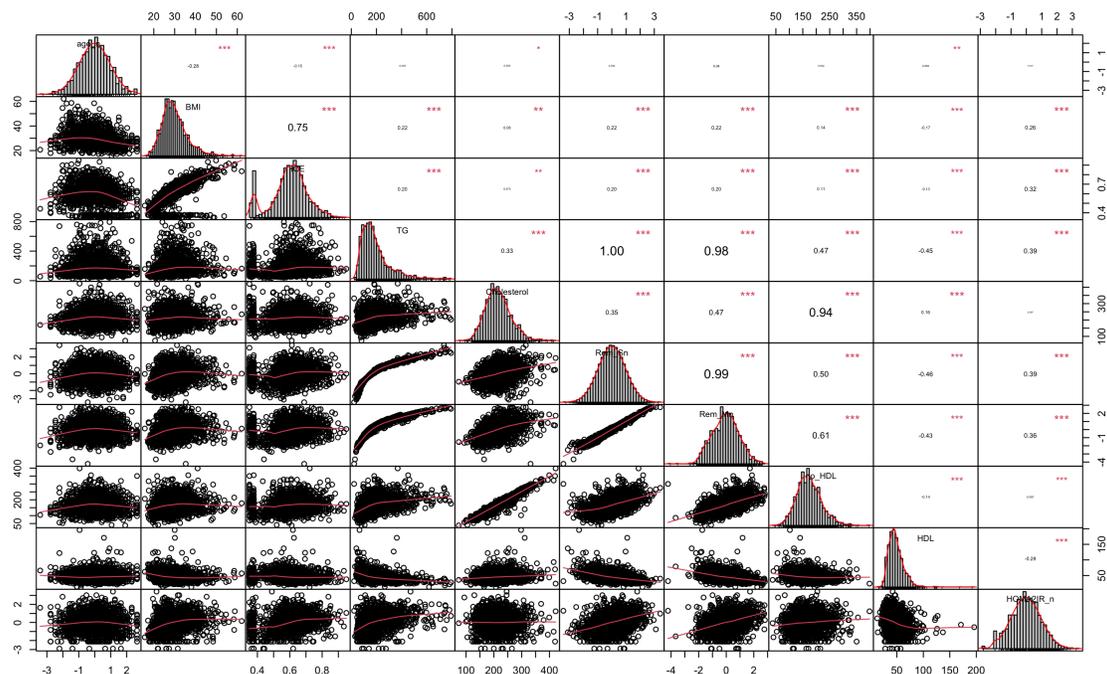


Figura 3.4: Gráficos descriptivos de las variables continuas normalizadas en sujetos con diabetes.

Entonces, de los 12,635 sujetos incluidos en el análisis, se eliminaron aquellos con triglicéridos mayores a 800mg/dL y niveles extremos de HOMA2IR ( $n = 48$ ), ya que niveles de triglicéridos en ese rango afecta la estimación (infraestimación) de las fórmulas utilizadas (Martin y Sampson). Por lo tanto, se analizaron en primera instancia a la submuestra de pacientes sin diabetes ( $n = 10, 990$ ). De igual forma se analizaron los casos con diabetes ( $n = 1, 597$ ) de forma independiente, *Figura 3.5*.

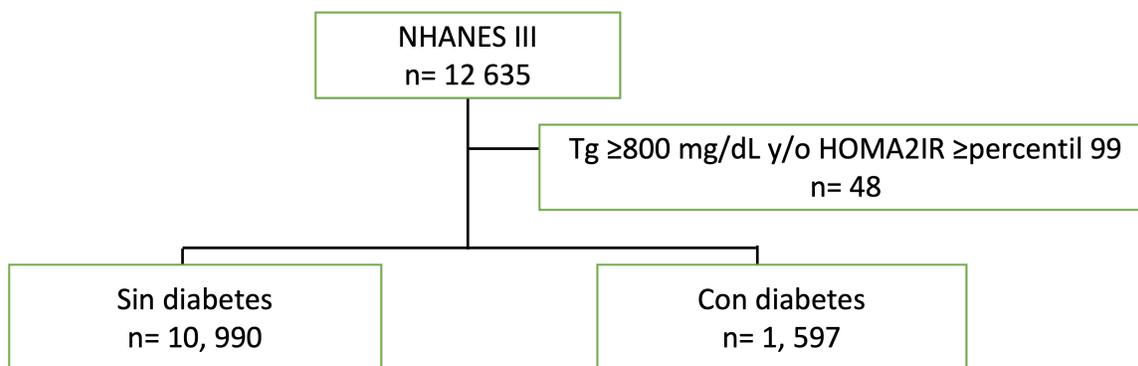


Figura 3.5: Flujograma de la organización de la muestra para el análisis.

## 2. Modelos de regresión lineal múltiple

Los modelos de RLM se ajustaron para evaluar la relación entre remanentes de colesterol y la resistencia de a la insulina, ajustado por edad, sexo, IMC, etnicidad, número de comorbilidades, tabaquismo y colesterol LDL. En cada modelo se analizó de forma gráfica y a través de pruebas estadísticas los supuestos del modelo. A continuación se describen los modelos en sujetos con diabetes y sin diabetes.

### 2.1. Modelos de regresión lineal múltiple en sujetos sin diabetes

#### 2.1.1. Modelo 1 de RLM para remanentes de colesterol (Sampson) en sujetos sin diabetes

En la *Figura 3.6* se muestra la relación entre remanentes de colesterol calculados con la fórmula de Sampson y la resistencia a la insulina evaluado con HOMA2IR.

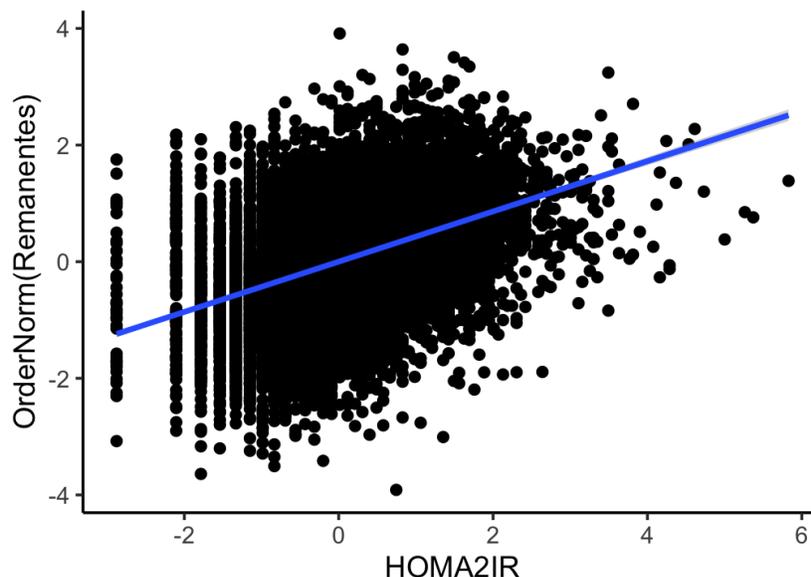


Figura 3.6: Gráficos de dispersión para evaluar la relación entre remanentes de colesterol (Sampson) y resistencia a la insulina (HOMA2IR) en sujetos sin diabetes.

Coeficientes	$\hat{\beta}$	Error estándar	Valor t	$p - value$
Intercepto	-0.781	0.035	-22.302	<0.0001
HOMA2IR	0.329	0.009	35.001	<0.0001
Edad	0.071	0.009	7.146	<0.0001
Sexo (Mujer)	-0.172	0.016	-10.522	<0.0001
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	0.546	0.022	24.268	<0.0001
Otro	0.406	0.045	9.104	<0.0001
Caucásico	0.407	0.021	19.590	<0.0001
IMC	0.135	0.009	14.575	<0.0001
Comorbilidades	0.0259	0.009	2.973	0.003
Tabaquismo	0.192	0.018	10.423	<0.0001
LDL-C	0.004	0.001	16.025	<0.0001
SRE= 0.840, 10917 G.L.	$R^2 = 0.294$	$R^2_{adj} = 0.294$	F-test=456.7	$p < 0.0001$

Tabla 3: Modelo 1 RLM saturado que incluye todas las variables estudiadas. Variable dependiente: Remanentes de colesterol-Sampson. Ref: categoría de referencia.

Como se observa en la **Tabla 2**, el modelo resultante tiene un  $R^2_{adj} = 0.294$  que indica que el modelo explica el 29.4% de la variabilidad de los niveles de remanentes de coleseterol. Dado el modelo se puede concluir que a mayor nivel de HOMA2IR mayores concentraciones de remanentes de colesterol;  $HOMA2IR \hat{\beta} = 0.329$ ,  $p$  –  $valor < 0.0001$ , independientemente de la edad, sexo, IMC, etnicidad, número de comorbilidades, tabaquismo y LDL-C. Posterior al ajuste del modelo, se evaluaron los supuestos.

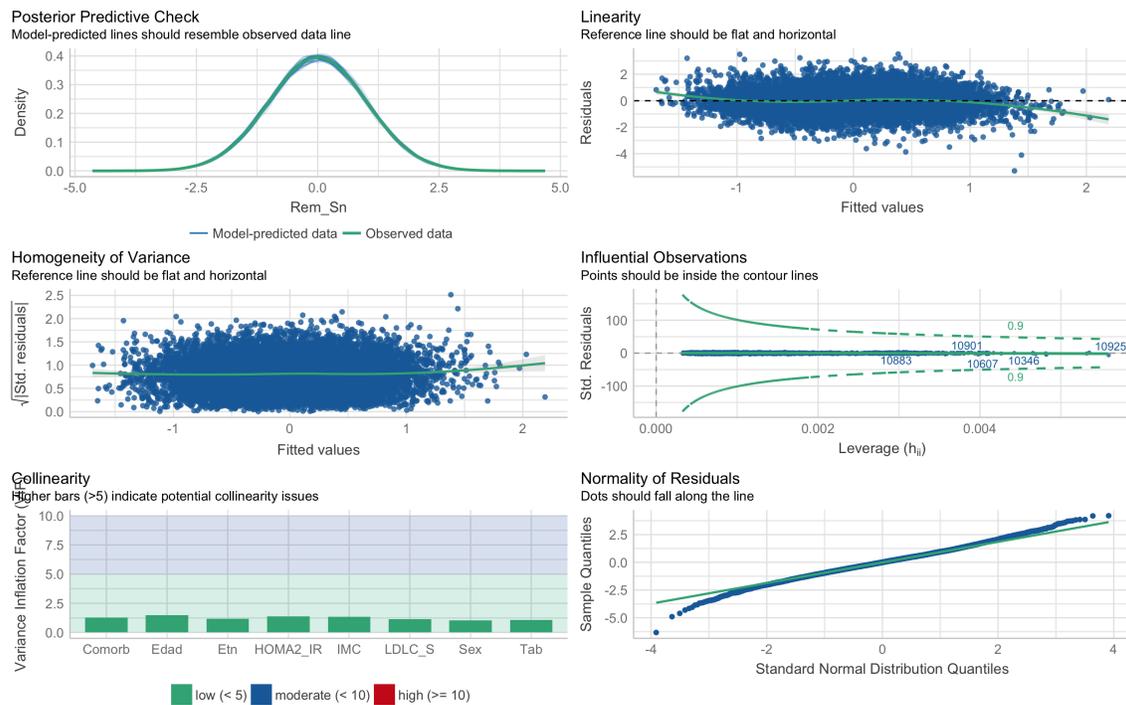


Figura 3.7: Evaluación de los supuestos del modelo 1 de RLM. Abreviaturas; Comorb: Número de comorbilidades; Etn: Etnicidad; IMC: Índice de masa corporal; Sex: Sexo; Tab: Tabaquismo.

Al evaluar el gráfico de linealidad de la *Figura 3.7*, se observa que la línea de ajuste es horizontal, aunque en los extremos se muestra cierta tendencia. Se puede concluir que el supuesto de linealidad se cumple. Para evaluar el supuesto de homoscedasticidad se realizó la prueba de Breusch-Pagan y dado el valor de p obtenido

( $p$ -valor = 0.0007) se podría concluir que no se cumple dicho supuesto. Sin embargo, al analizar el gráfico de la *Figura 3.7* no se observa tendencia alguna con respecto a los valores ajustado y la raíz cuadrada de los residuales estandarizados. Por lo tanto, con la inspección visual de dicho gráfico se puede concluir que el supuesto de homoscedasticidad se cumple. Probablemente, debido al tamaño de muestra y que la prueba de Breusch-Pagan es una prueba sensible, se estén detectando variaciones pequeñas en la relación de los valores ajustado por el modelo y la raíz cuadrada de los residuales estandarizados. Además, se realizó la prueba de Breusch-Godfrey para evaluar el supuesto de no autocorrelación de los errores y dado el  $p$ -valor = 0.971 se concluye que los errores no están correlacionados. Para evaluar la distribución de los errores (residuales) se ajustaron las prueba de Jarque-Bera y Anderson-Darling que indican que los errores no siguen una distribución normal ( $p$ -valor < 0.0001 en ambos casos). Sin embargo, al analizar el gráfico Q-Q de la *Figura 3.7* se observa que los errores se ajustan a una distribución normal estándar, pero con mayor dispersión de los datos en las colas de la distribución, lo que podría estar detectando la prueba de Anderson-Darling: a pesar de esto, no esperamos que esta desviación altere de forma significativa las inferencias obtenidas a partir del modelo. Por ende, se concluye que los errores siguen aproximadamente una distribución normal.

Intervalos de confianza simultáneos al 95 %			
Variable	Estimador	Límite inferior	Límite superior
Intercepto	-0.781	-0.879	-0.682
HOMA2IR	0.329	0.302	0.355
Edad	0.071	0.041	0.098
Sexo (Mujer)	-0.172	-0.217	-0.125
Afroamericanos	Ref	Ref	Ref
Mexicoamericanos	0.546	0.483	0.609
Otros	0.406	0.280	0.530
Caucásicos	0.407	0.349	0.465
IMC	0.136	0.109	0.162
Comorbilidades	0.027	0.001	0.052
Tabaquismo	0.192	0.140	0.244
LDL-C	0.004	0.003	0.005

Tabla 4: Intervalos de confianza simultáneos al 95 % para los coeficientes del modelo 1 de RLM (Sampson).

Por último, se analizó la influencia de posibles valores atípicos en el gráfico de la distancia de Cook y se observa que ningún dato está fuera de la línea punteada, entonces no se identificaron posibles valores atípicos. Para evaluar el supuesto de colinealidad, se calculó el factor de inflación de la varianza (VIF) y, tal y como se presenta en la gráfica, todas las covariables incluidas en el modelo tuvieron un VIF cercano a 1, lo que indicaría que no hay colinealidad o ésta es muy pequeña. Finalmente, se calcularon los intervalos de confianza simultáneos al 95 % para el intercepto y las covariables incluidas en el modelo, *Tabla 4*. Para el modelo 1, se observó que tanto los intervalos del intercepto como de las covariables fueron estadísticamente significativos. Por lo tanto, se concluye el modelo 1 de RLM ajusta adecuadamente los datos.

### 2.1.2. Modelo 2 de RLM para remanentes de colesterol (Martin) en sujetos sin diabetes

En la *Figura 3.8* se muestra la relación entre remanentes de colesterol calculados con la fórmula de Martin (Rem-M) y la resistencia a la insulina evaluada con HOMA2IR.

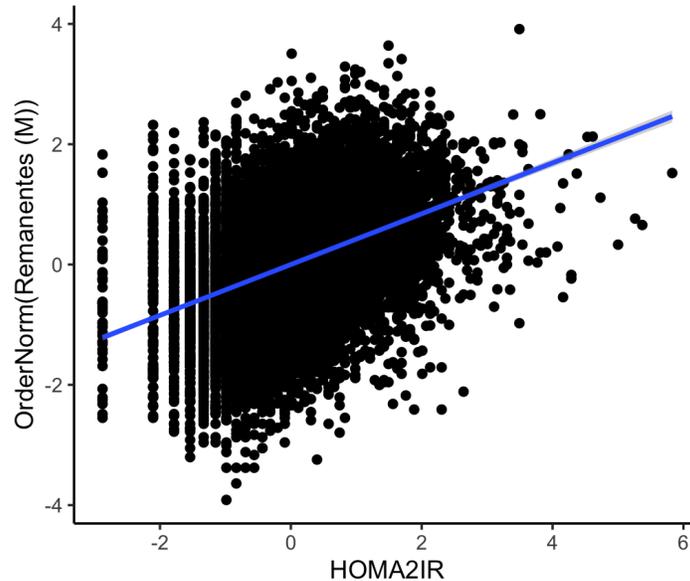


Figura 3.8: Gráficos de dispersión para evaluar la relación entre remanentes de colesterol (Martin) y resistencia a la insulina (HOMA2IR) en sujetos sin diabetes.

Como se observa en la **Tabla 5**, el modelo resultante tiene un  $R_{adj}^2 = 0.412$  que indica que el modelo explica el 41.2 % de la variabilidad de los niveles de remanentes de colesterol. Considerando el modelo se puede concluir que a mayor nivel de HOMA2IR mayores concentraciones de remanentes de colesterol;  $HOMA2IR \hat{\beta} = 0.293$ ,  $p - valor < 0.001$ , independientemente de la edad, sexo, IMC, etnicidad, número de comorbilidades, tabaquismo y LDL-C. Posterior al ajuste del modelo, se evaluaron los supuestos, tal como se presenta en la *Figura 3.9*.

Coefficientes	$\hat{\beta}$	Error estándar	Valor t	$p - value$
Intercepto	-1.559	0.032	-48.702	<0.001
HOMA2IR	0.293	0.009	34.162	<0.001
Edad	0.059	0.009	6.548	<0.001
Sexo (Mujer)	-0.147	0.015	-9.899	<0.001
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	0.475	0.021	23.085	<0.001
Otro	0.332	0.041	8.165	<0.001
Caucásico	0.337	0.019	17.792	<0.001
IMC	0.110	0.009	12.968	<0.001
Comorbilidades	0.023	0.008	2.815	0.005
Tabaquismo	0.175	0.017	10.375	<0.001
LDL-C	0.010	0.001	48.721	<0.001
SRE= 0.767, 10917 G.L.	$R^2 = 0.413$	$R^2_{adj} = 0.412$	F-test=767.6	p<0.001

Tabla 5: Modelo 2 RLM saturado que incluye todas las variables estudiadas. Variable dependiente: Remanentes de colesterol-Martin (Rem-M). Ref: categoría de referencia.

Al evaluar el gráfico de linealidad de la *Figura 3.9*, se observa que la línea de ajuste es horizontal, aunque en los extremos se muestra cierta tendencia. Sin embargo, se concluye que el supuesto de linealidad se cumple. Para evaluar el supuesto de homoscedasticidad se ajustó la prueba de Breusch-Pagan y considerando el valor de p obtenido ( $p - valor = 0.0059$ ) se podría concluir que no se cumple dicho supuesto. Pero, al analizar el gráfico de la *Figura 3.9* no se observa tendencia alguna con respecto a los valores ajustado y la raíz cuadrada de los residuales estandarizados. Por lo tanto, con la inspección visual de dicho gráfico se puede concluir que el supuesto de homoscedasticidad se cumple. Probablemente, debido al tamaño de muestras y que la prueba de Breusch-Pagan es una prueba sensible, se estén detectando variaciones pequeñas. Se realizó la prueba de Breusch-Godfrey para evaluar el supuesto de no

autocorrelación de los errores y dado el  $p$ -valor = 0.8972 se concluye que los errores no están correlacionados. Para evaluar la distribución de los errores se ajustaron las prueba de Anderson-Darling y Jarque-Bera indicando que los errores no siguen una distribución normal ( $p$ -valor < 0.0001 en ambos casos). Sin embargo, al analizar el gráfico Q-Q de la *Figura 3.9* se observa que los errores se ajustan a una distribución normal estándar, pero con una discreta dispersión de los datos en las colas de la distribución, lo que podría estar detectando la prueba de Anderson-Darling y Jarque-Bera. Por ende, se concluye que los errores siguen aproximadamente una distribución normal.

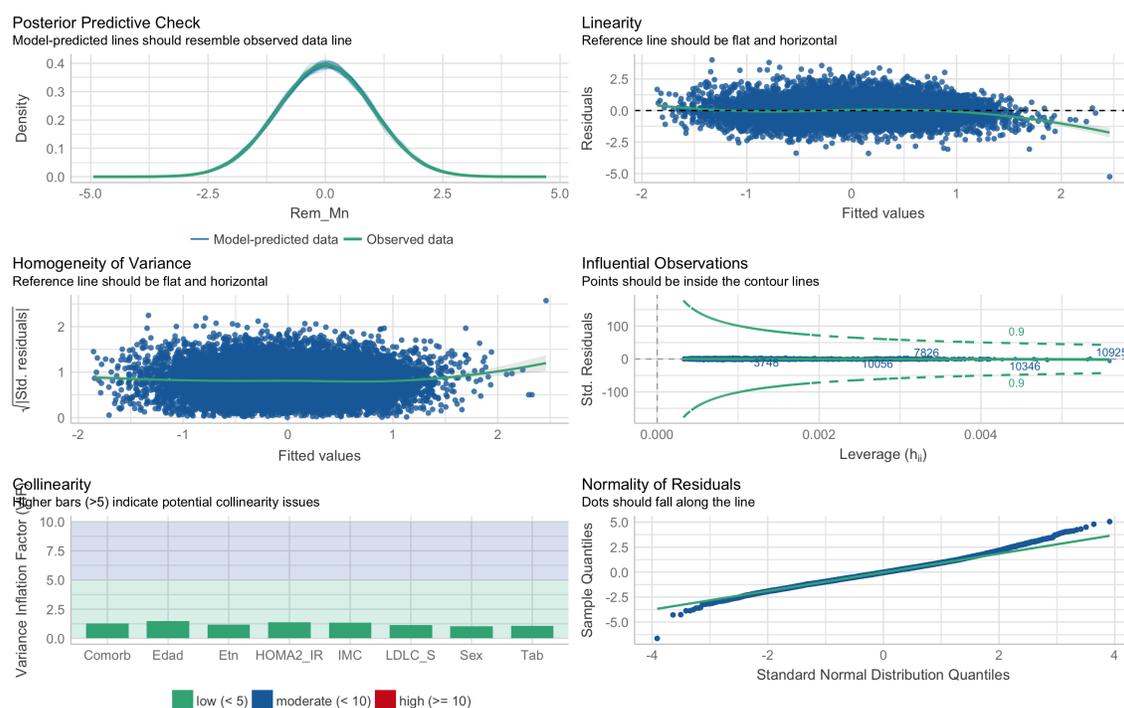


Figura 3.9: Evaluación de los supuestos del modelo 2 de RLM. Abreviaturas; Comorb: Número de comorbilidades; Etn: Etnicidad; IMC: Índice de masa corporal; Sex: Sexo; Tab: Tabaquismo.

Para evaluar la influencia de posibles valores atípicos, se analizó el gráfico de la distancia de Cook y se observó que ningún dato está fuera de la línea punteada, entonces no se identificaron posibles valores atípicos. Considerando el VIF, tal y como se presenta en la gráfica, todas las covariables incluidas en el modelo tuvieron un VIF

cercano a 1, lo que indicaría que no hay colinealidad o ésta es muy pequeña. Finalmente, se calcularon los intervalos de confianza simultáneos al 95 % para el intercepto y las covariables incluidas en el modelo, *Tabla 6*. Para el modelo 2, se observó que tanto los intervalos del intercepto como de las covariables fueron estadísticamente significativas. Por lo tanto, se concluye el modelo 2 de RLM ajusta adecuadamente los datos.

Intervalos de confianza simultáneos al 95 %			
Variable	Estimador	Límite inferior	Límite superior
Intercepto	-1.560	-1.649	-1.470
HOMA2IR	0.293	0.269	0.317
Edad	0.059	0.033	0.084
Sexo (Mujer)	-0.147	-0.189	-0.105
Afroamericanos	Ref	Ref	Ref
Mexicoamericanos	0.475	0.416	0.532
Otros	0.332	0.217	0.446
Caucásicos	0.337	0.285	0.391
IMC	0.110	0.086	0.134
Comorbilidades	0.023	0.004	0.046
Tabaquismo	0.175	0.127	0.218
LDL-C	0.010	0.009	0.011

Tabla 6: Intervalos de confianza simultáneos al 95 % para los coeficientes del modelo 2 de RLM (Martin).

## 2.2. Modelos de regresión lineal múltiple en sujetos con diabetes

### 2.2.1. Modelo 3 de RLM para remanentes de colesterol (Sampson) en sujetos con diabetes

En la *Figura 3.10* se muestra la relación entre remanentes de colesterol calculados con la fórmula de Sampson (Rem-S) y la resistencia a la insulina evaluada con HOMA2IR en sujetos con diabetes.

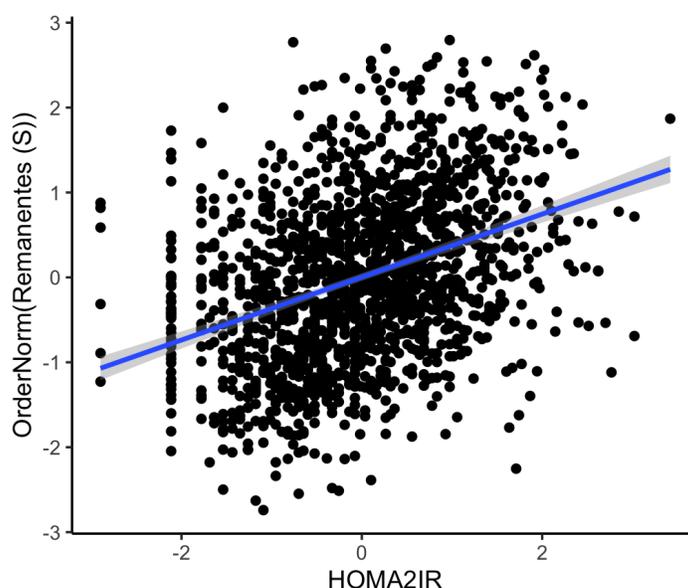


Figura 3.10: Gráfico de dispersión para evaluar la relación entre remanentes de colesterol (Sampson) y resistencia a la insulina (HOMA2IR) en sujetos con diabetes.

El modelo resultante tiene un  $R_{adj}^2 = 0.212$  que indica que el modelo explica el 21.2% de la variabilidad de los niveles de remanentes de colesterol. Considerando el modelo se puede concluir que a mayor nivel de HOMA2IR mayores concentraciones de remanentes de colesterol en sujetos con diabetes;  $HOMA2IR \hat{\beta} = 0.343$ ,  $p - valor < 0.0001$ , independientemente de la edad, sexo, IMC, etnicidad, número de comorbilidades, tabaquismo y LDL-C. Posterior al ajuste del modelo, se evaluaron los supuestos, tal como se presenta en la *Figura 3.11*.

Coefficientes	$\hat{\beta}$	Error estándar	Valor t	$p - value$
Intercepto	-1.304	0.142	-9.179	<0.0001
HOMA2IR	0.343	0.023	14.891	<0.001
Edad	-0.007	0.025	-0.254	0.799
Sexo (Mujer)	-0.018	0.045	-0.416	0.677
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	0.474	0.058	8.169	<0.001
Otro	0.188	0.126	1.488	0.137
Caucásico	0.394	0.057	6.957	<0.001
IMC	0.021	0.003	5.458	0.0001
Comorbilidades	0.034	0.016	2.125	0.034
Tabaquismo	0.006	0.056	0.118	0.906
LDL-C	0.003	0.0005	4.768	<0.001
SRE= 0.888, 1641 G.L.	$R^2 = 0.217$	$R_{adj}^2 = 0.212$	F-test=45.45	$p < 0.001$

Tabla 7: Modelo 3 RLM saturado que incluye todas las variables estudiadas. Variable dependiente: Remanentes de colesterol-Sampson (Rem-S). Ref: categoría de referencia.

Al evaluar el gráfico de linealidad de la *Figura 3.11*, se observa que la línea de ajuste es horizontal, aunque en los extremos se muestra cierta tendencia e intervalos más amplios. Sin embargo, se concluye que el supuesto de linealidad se cumple. Para evaluar el supuesto de homoscedasticidad se ajustó la prueba de Breusch-Pagan y considerando el valor de  $p$  obtenido ( $p - valor = 0.821$ ), además del gráfico presentado en la *Figura 3.11* se concluye que se cumple dicho supuesto. Se realizó la prueba de Breusch-Godfrey para evaluar el supuesto de no autocorrelación de los errores y dado el  $p - valor = 0.427$  se concluye que los errores no están correlacionados. Para evaluar la distribución de los errores se ajustó la prueba de Anderson-Darling y dado el  $p - valor = 0.114$  se concluye que la distribución residuales se ajustan a una dis-

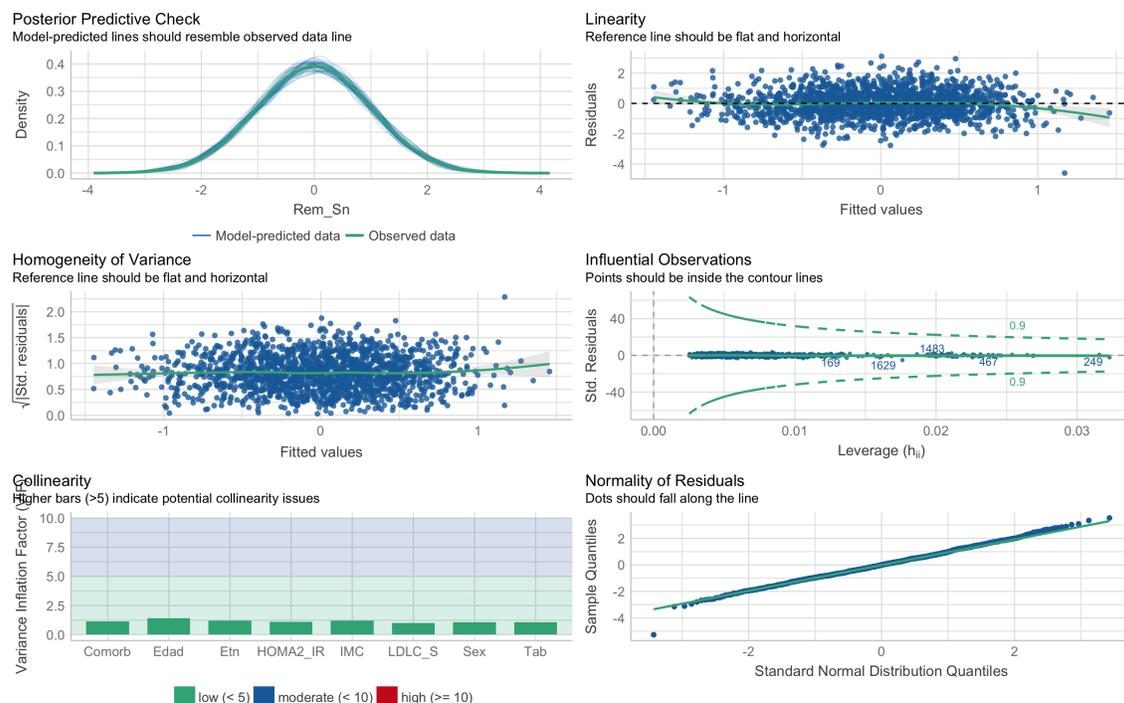


Figura 3.11: Evaluación de los supuestos del modelo 3 de RLM. Abreviaturas; Co-morb: Número de comorbilidades; Etn: Etnicidad; IMC: Índice de masa corporal; Sex: Sexo; Tab: Tabaquismo.

tribución normal estándar. Además, al analizar el gráfico Q-Q de la *Figura 3.11* se observa que los errores se ajustan a una distribución normal estándar, pero con una discreta dispersión de los datos en las colas de la distribución. Por ende, se concluye que los errores siguen aproximadamente una distribución normal.

Además, se evaluó la influencia de posibles valores atípicos, al analizar el gráfico de la distancia de Cook se observa que ningún dato está fuera de la línea punteada, entonces puede concluir que no hay valores atípicos que estén afectando la estimación del modelo. Considerando el VIF, tal y como se presenta en la gráfica, todas las covariables incluidas en el modelo tuvieron un VIF cercano a 1, lo que indicaría que no hay colinealidad o ésta es muy pequeña. Finalmente, se calcularon los intervalos de confianza simultáneos al 95% para el intercepto y las covariables incluidas en el modelo, *Tabla 6*. Para el modelo 2, se observó que los intervalos del modelo para las variables de interés, etnicidad, IMC, número de comorbilidades y LDL-C fueron

estadísticamente significativas, en cambio para la edad, sexo y tabaquismo no fueron estadísticamente significativos, coincidiendo con el valor de  $p$  del modelo 3 de RLM ajustado previamente. En resumen, se concluye el modelo 3 de RLM ajusta adecuadamente los datos.

Intervalos de confianza simultáneos al 95 %			
Variable	Estimador	Límite inferior	Límite superior
Intercepto	-1.304	-1.705	-0.904
HOMA2IR	0.343	0.278	0.408
Edad	-0.007	-0.079	0.066
Sexo (Mujer)	-0.018	-0.146	0.109
Afroamericanos	Ref	Ref	Ref
Mexicoamericanos	0.475	0.311	0.638
Otros	0.188	-0.168	0.545
Caucásicos	0.394	0.234	0.554
IMC	0.021	0.009	0.031
Comorbilidades	0.034	-0.011	0.079
Tabaquismo	0.006	-0.150	0.164
LDL-C	0.003	0.001	0.004

Tabla 8: Intervalos de confianza simultáneos al 95 % para los coeficientes del modelo 3 de RLM (Sampson). Ref: categoría de referencia.

### 2.2.2. Modelo 4 de RLM para remanentes de colesterol (Martin) en sujetos con diabetes

En la *Figura 3.12* se muestra la relación entre remanentes de colesterol calculados con la fórmula de Martin y la resistencia a la insulina evaluado con HOMA2IR en sujetos con diabetes.

El modelo resultante tiene un  $R_{adj}^2 = 0.2976$  que indica que el modelo explica el 29.76 % de la variabilidad de los niveles de remanentes de coleseterol. Considerando el modelo se puede concluir que a mayor nivel de HOMA2IR mayores concentraciones

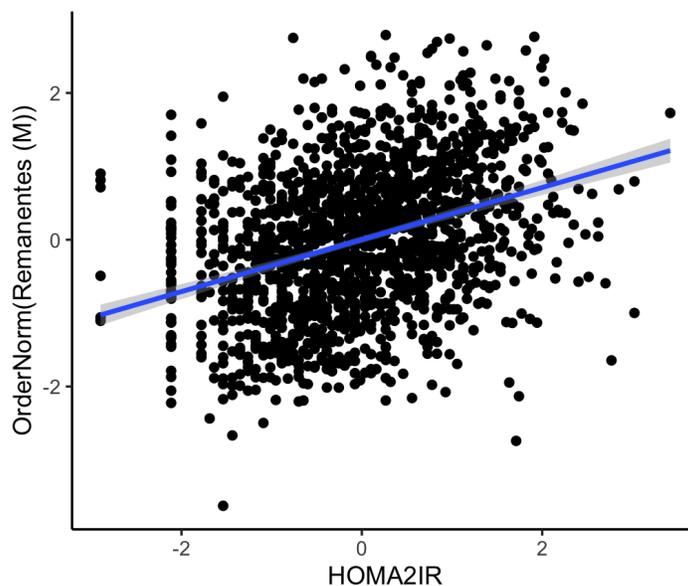


Figura 3.12: Gráficos de dispersión para evaluar la relación entre remanentes de colesterol (Martin) y resistencia a la insulina (HOMA2IR) en sujetos con diabetes. Ref: categoría de referencia.

de remanentes de colesterol en sujetos con diabetes;  $HOMA2IR \hat{\beta} = 0.319$ ,  $p - valor < 0.0001$ , independientemente de la edad, sexo, IMC, etnicidad, número de comorbilidades, tabaquismo y LDL-C, *Tabla 9*.

En seguida, se procedió a evaluar los supuestos del modelo, *Figura 3.13*. Al evaluar el gráfico de linealidad, se observa que la línea de ajuste es horizontal, por lo tanto se concluye que el supuesto de linealidad se cumple. Para evaluar el supuesto de homoscedasticidad se ajustó la prueba de Breusch-Pagan y considerando el valor de  $p$  obtenido ( $p - valor = 0.188$ ), además del gráfico presentado en la *Figura 3.11* se concluye que se cumple dicho supuesto. Se realizó la prueba de Breusch-Godfrey para evaluar el supuesto de no autocorrelación de los errores y dado el  $p - valor = 0.093$  se concluye que los errores no están correlacionados. Para evaluar la distribución de los errores se ajustó la prueba de Anderson-Darling y si consideramos la prueba se concluiría que los errores no siguen una distribución normal ( $p - valor = 0.04316$ ). Sin embargo, al analizar el gráfico Q-Q de la *Figura 3.11* se observa que los errores se ajustan a una distribución normal estándar, pero con una discreta dispersión de los datos en las colas de la distribución, lo que podría influir en el resultado de la prueba

de Anderson-Darling. Por ende, se concluye que los errores siguen aproximadamente una distribución normal.

Coeficientes	$\hat{\beta}$	Error estándar	Valor t	$p - value$
Intercepto	-2.082	0.169	-12.272	<0.001
HOMA2IR	0.319	0.022	14.695	<0.001
Edad	0.001	0.002	0.615	0.533
Sexo (Mujer)	-0.009	0.002	-0.218	0.827
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	0.463	0.055	8.438	<0.001
Otro	0.182	0.120	1.522	0.128
Caucásico	0.369	0.053	6.938	<0.001
IMC	0.019	0.003	5.120	<0.001
Comorbilidades	0.027	0.015	1.778	0.075
Tabaquismo	0.022	0.052	0.411	0.681
LDL-C	0.008	0.0005	16.090	<0.001
SRE= 0.838, 1641 G.L.	$R^2 = 0.302$	$R^2_{adj} = 0.2976$	F-test=70.94	p<0.001

Tabla 9: Modelo 4 RLM saturado que incluye todas las variables estudiadas. Variable dependiente: Remanentes de colesterol-Martin.

Por último, se evaluó la influencia de posibles valores atípicos, al analizar el gráfico de la distancia de Cook se observa que ningún dato está fuera de la línea punteada, entonces puede concluir que no hay valores atípicos que estén afectando la estimación del modelo. Considerando el VIF, tal y como se presenta en la gráfica, todas las covariables incluidas en el modelo tuvieron un VIF cercano a 1, lo que indicaría que no hay colinealidad o ésta es muy pequeña.

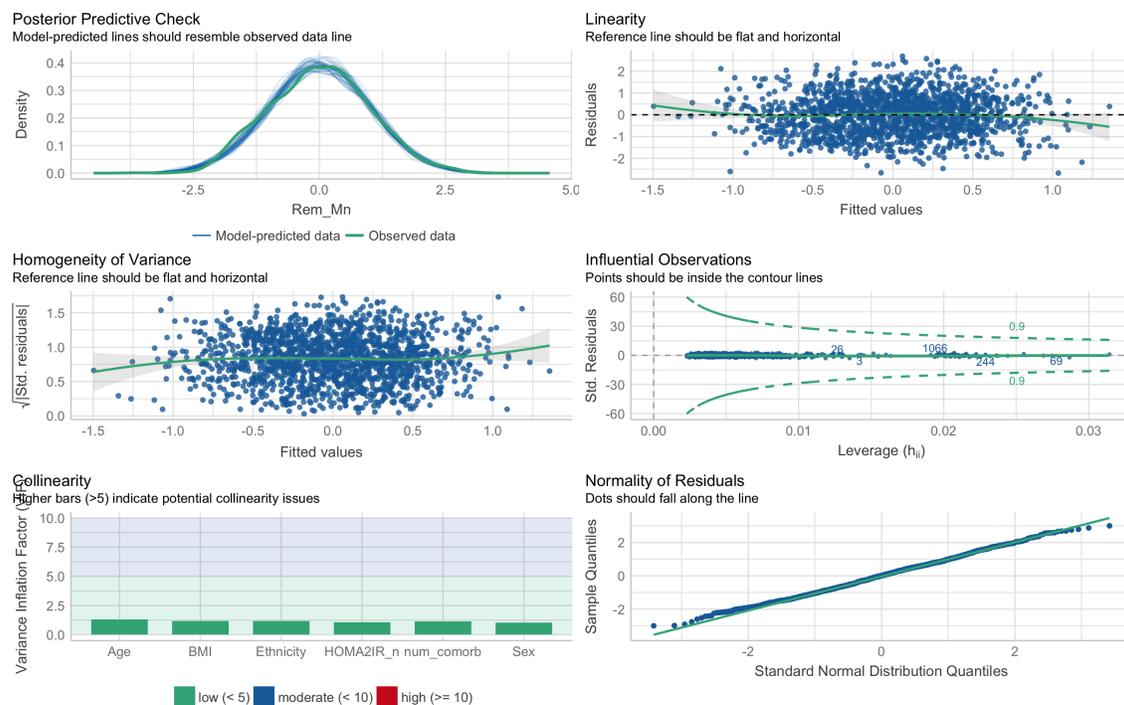


Figura 3.13: Evaluación de los supuestos del modelo 4 de RLM. Abreviaturas; Co-morb: Número de comorbilidades; Etn: Etnicidad; IMC: Índice de masa corporal; Sex: Sexo; Tab: Tabaquismo.

Intervalos de confianza simultáneos al 95 %			
Variable	Estimador	Límite inferior	Límite superior
Intercepto	-2.082	-2.561	-1.604
HOMA2IR	0.319	0.258	0.380
Edad	0.001	-0.003	0.005
Sexo (Mujer)	-0.009	-0.129	0.110
Afroamericanos	Ref	Ref	Ref
Mexicoamericanos	0.463	0.308	0.617
Otros	0.182	-0.155	0.519
Caucásicos	0.369	0.219	0.520
IMC	0.019	0.008	0.028
Comorbilidades	0.027	-0.016	0.069
Tabaquismo	0.022	-0.126	0.169
LDL-C	0.008	0.006	0.009

Tabla 10: Intervalos de confianza simultáneos al 95 % para los coeficientes del modelo 4 de RLM (Martin). Ref: categoría de referencia.

Finalmente, se calcularon los intervalos de confianza simultáneos al 95 % para el intercepto y las covariables incluidas en el modelo, *Tabla 6*. Para el modelo 4, se observó que los intervalos del modelo para las variables de interés, etnicidad, IMC y LDL-C fueron estadísticamente significativas. En cambio para la edad, sexo, número de comorbilidades y tabaquismo no fueron estadísticamente significativos, coincidiendo con el valor de p del modelo 4 de RLM ajustado previamente. En resumen, se concluye el modelo 4 de RLM ajusta adecuadamente los datos.

### 3. Modelos de riesgos competitivos de Fine y Gray

Para evaluar el riesgo de mortalidad cardiovascular que confiere la resistencia a la insulina (HOMA2IR) y los remanentes de colesterol en sujetos con y sin diabetes, se ajustaron modelos de riesgos competitivos de Fine y Gray. Se consideró este abordaje de análisis de supervivencia porque la mortalidad por otra causa distinta a la cardiovascular influiría en la estimación del riesgo asociado a la resistencia a la insulina y los remanentes de colesterol.

Para facilitar la interpretación de los sHR de los modelos de Fine y Gray, se categorizó la variable HOMA2IR acorde al percentil 85, en cambio para los remanentes de colesterol la variable se centró y estandarizó. Por lo tanto, las variables HOMA2IR, sexo, tabaquismo y etnicidad se trataron como variables cualitativas dentro del análisis. Por otro lado, edad, IMC, remanentes de colesterol, LDL-C y número de comorbilidades se incluyeron como variables cuantitativas respetando la distribución que tomaron dichas variables en la muestra. En los siguientes apartados se describirán los modelos de riesgos competitivos de Fine y Gray.

### 3.1. Modelos de riesgos competitivos de Fine y Gray en sujetos sin diabetes

Como se describió en la *Tabla 2*, las tasas de incidencia fueron mayores en la submuestra de sujetos con diabetes. Por ende, se determinó analizar el riesgo que confiere la resistencia a la insulina (HOMA2IR) y los remanentes de colesterol de forma independiente acorde a la presencia de diabetes o no. Además, debido a la diabetes, era esperado niveles más altos de remanentes de colesterol dada la hipertrigliceridemia observada en pacientes con diabetes y, debido a la fisiopatología es esperado que el valor de HOMA2IR sea mucho más alto en sujetos con diabetes. Entonces, esto podría influir en la estimación del riesgo de subdistribución asociado a dichas variables. A continuación se presentan los modelos en sujetos sin diabetes.

#### 3.1.1. Modelos 1 de riesgos competitivos de Fine y Gray en sujetos sin diabetes

En la *Tabla 11* se muestra el modelo de Fine y Gray para evaluar el riesgo de mortalidad cardiovascular. Con base en los resultados, se puede decir que por cada aumento en 1 desviación estándar en la concentración de remanentes de colesterol estimados con la fórmula de Sampson aumenta en un 32.53 % la incidencia relativa de muerte a causa de un evento cardiovascular. En cambio, un  $HOMA2IR \geq 2.5$  aumentó en un 19.72 % la incidencia relativa de muerte a causa de un evento cardiovascular independientemente de la edad, sexo, IMC, etnicidad, tabaquismo, LDL-C y número de comorbilidades. Además, se probó la interacción en el modelo de  $HOMA2IR * Rem - S$ , pero no fue estadísticamente significativa,  $p - valor = 0.488$ , lo que indica que no existe un efecto modificador entre las variables y, por tanto, es factible explorar un posible efecto mediador. Sin embargo, en el modelo se incluyó la interacción de remanentes de colesterol y edad. Interesantemente, la edad modifica el efecto de los remanentes de colesterol sobre la mortalidad cardiovascular. Por último, se evaluó el  $C - statistic$  o concordancia una medida de bondad de ajuste utilizada en modelos de supervivencia que indica la probabilidad de que la predicción  $x$  vaya en la misma

dirección que los datos reales  $y$ . Por lo tanto, un par de observaciones  $i, j$  se considera concordante si la predicción y los datos van en la misma dirección. La concordancia o  $C - statistic$ , una medida de discriminación del modelo donde un valor mayor a 0.7 indica un modelo que discrimina adecuadamente a los sujetos que presentan el evento de quienes no. En este caso, el  $C - statistic = 0.850$  que indica un modelo que tiene una buena discriminación.

Variable	$\hat{\beta}$	sHR	IC95 %	$p - value$
$Rem - S_+$	0.281	1.325	1.060-1.658	0.0137
$HOMA2IR \geq 2.5$	0.180	1.197	1.038-1.380	0.0133
Edad	0.068	1.070	1.066-1.075	<0.0001
Sexo (Mujer)	-0.250	0.778	0.0.690-0.877	<0.0001
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	-0.039	0.961	0.792-1.166	0.6901
Otro	-0.124	0.883	0.611-1.276	0.5081
Caucásico	-0.199	0.819	0.701-0.957	0.0122
IMC	0.012	1.012	0.992-1.025	0.0656
Comorbilidades	0.139	1.149	1.091-1.210	<0.0001
Tabaquismo	0.358	1.431	1.240-1.652	<0.0001
$LDL - C_+$	0.127	1.135	1.071-1.204	<0.0001
$Edad/Rem - S^*$	-0.004	0.996	0.992-0.999	0.0160
n=146458, eventos=1166	$C - statistic = 0.850$	$SE = 0.005$	BIC=19590.76	

Tabla 11: Modelo 1 de riesgos competitivos de Fine y Gray para evaluar mortalidad cardiovascular. Rem-S: Remanentes de colesterol-Sampson. sHR: subdistribution Hazard ratio. \* Interacción; +los valores fueron centrados y estandarizados.

Posteriormente, se analizó a detalle la interacción entre remanentes de colesterol y la edad de los sujetos. Para ello, se utilizó el paquete *simPH* para simular y graficar

la interacción entre dichas variables. En la *Figura 3.14*, se observa que en pacientes menores de 70 años a mayor concentración de remanentes de colesterol calculados se asocian con una mayor incidencia relativa de muerte a causa de un evento cardiovascular. En cambio, en sujetos con una edad igual o mayor a 70 años, una concentración mayor de remanentes de colesterol se asocia con una disminución discreta de la incidencia relativa de muerte a causa de un evento cardiovascular; aunque dichos sujetos tienen mayor riesgo de mortalidad cardiovascular en comparación con sujetos menores de 70 años.

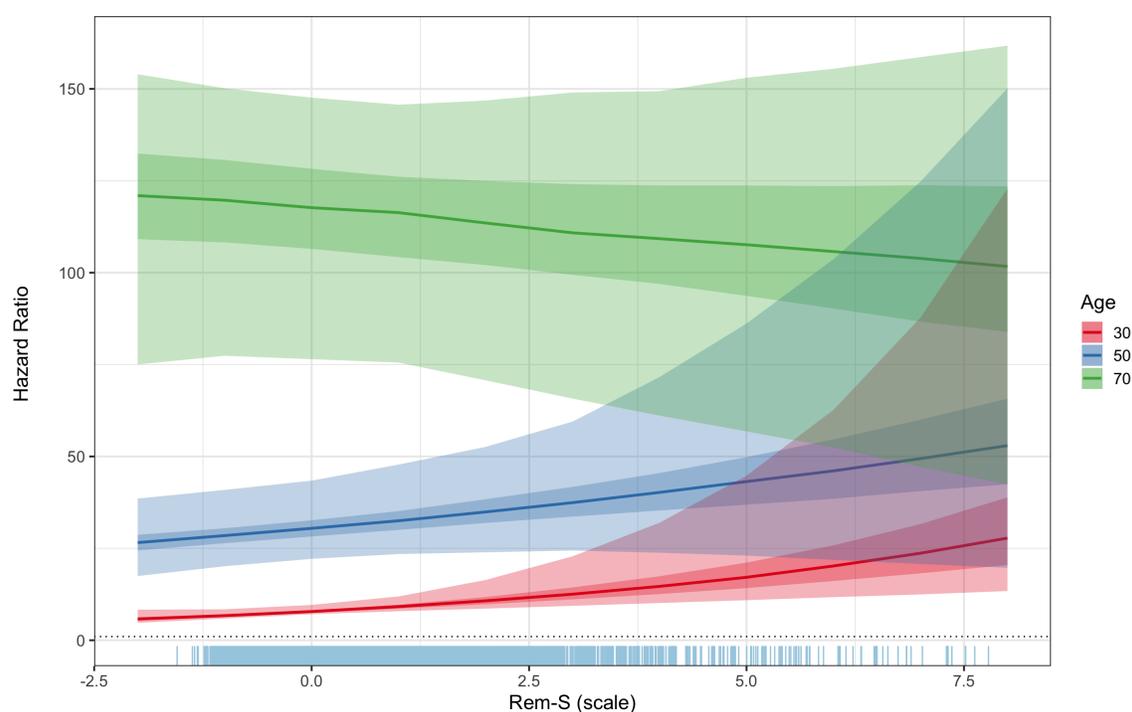


Figura 3.14: Evaluación de la interacción entre edad y remanentes de colesterol sobre la mortalidad cardiovascular en sujetos sin diabetes. Rangos de edad: 30 – 49, 50 – 69 y  $\geq 70$  años.

Para evaluar el ajuste del modelo se analizaron de forma visual los residuos de Schoenfeld, *Figura 3.15*. Como se puede observar, la distribución de los residuos para cada variable independiente incluida en el modelo no muestra una tendencia clara. El supuesto de riesgos proporcionales indica que los residuos son independientes del tiempo. Por lo tanto, una gráfica que presenta un patrón no aleatorio contra el tiempo será evidencia de la violación del supuesto.

Sin embargo, al aplicar la función `cox.zph()` que para cada variable correlaciona el conjunto correspondiente de residuales estandarizados con el tiempo, para probar la independencia entre ambos. Muestra que para el caso de HOMA2IR y remanentes de colesterol, el  $p$  – *valor* es mayor a 0.05; 0.664 y 0.1422, respectivamente. Lo que indica independencia de los residuales y el tiempo, pero el resto de variables y el análisis global tiene un  $p$  – *valor*  $\leq 0.001$  lo que indica que hay una correlación entre el tiempo y los residuales. Sin embargo, el tamaño de muestra y que el modelo de riesgos competitivos de Fine y Gray es una adaptación del modelo de riesgos proporcionales de Cox podría estar influyendo el resultado de la prueba, a pesar de que visualmente no se observa una tendencia clara. Dada estas limitantes, se concluye que el modelo ajusta adecuadamente los datos.

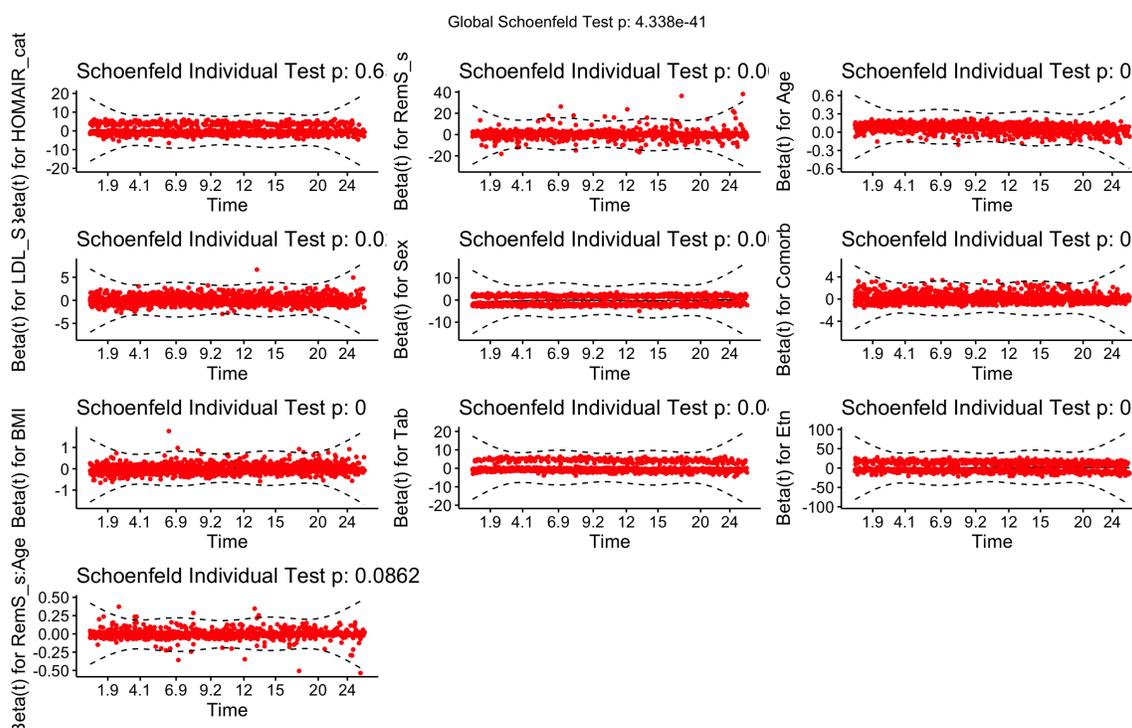


Figura 3.15: Supuesto de riesgos proporcionales: residuos de Schoenfeld.

Para evaluar posibles observaciones influyentes, se analizaron los valores  $dfbeta$  para cada coeficiente utilizando la función `ggcoxdiagnostics()`, *Figura 16*. El argumento `type="dfbeta"` de la función previa traza los cambios estimados en los coeficientes de

regresión al eliminar cada observación a su vez.

Al analizar los gráficos se observa que no existen valores realmente influyentes, los coeficientes de cada variable cambian muy poco al eliminar ciertas observaciones. Por lo tanto, no es necesario limitar aún más la muestra.

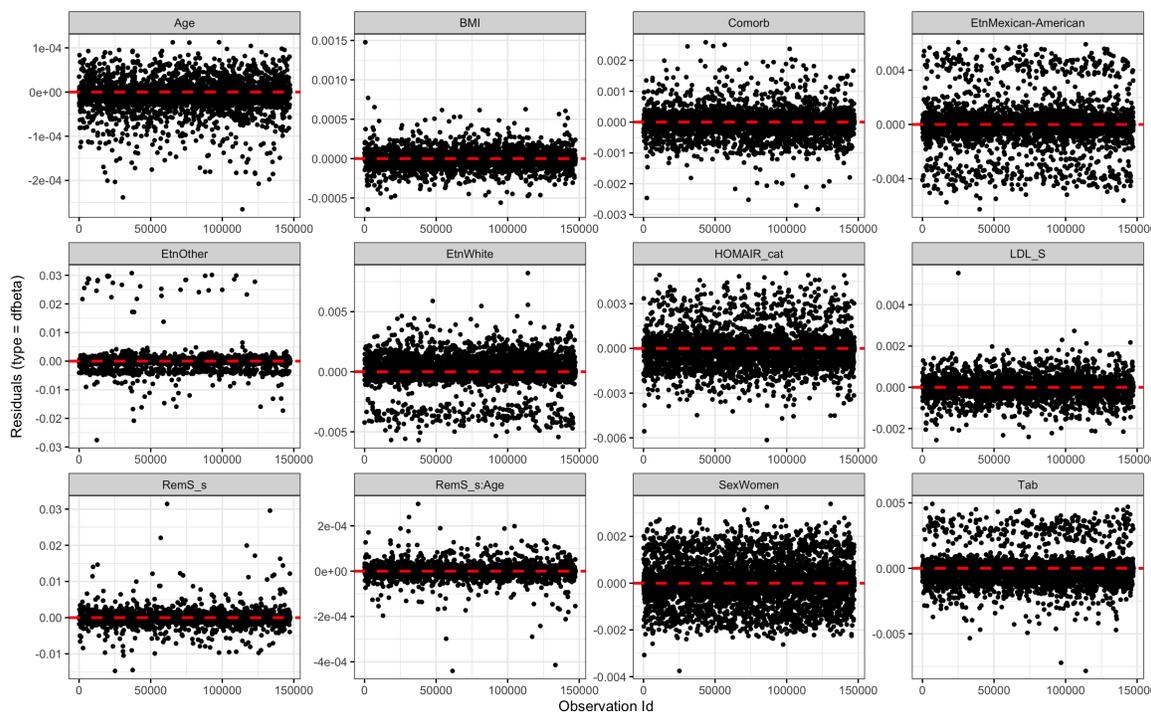


Figura 3.16: Gráficos de residuales (*dfbeta*) para evaluar posibles observaciones influyentes en el modelo de riesgos competitivos.

### 3.1.2. Modelos 2 de riesgos competitivos de Fine y Gray en sujetos sin diabetes

En la *Tabla 12* se muestra el modelo de Fine y Gray para evaluar el riesgo de mortalidad cardiovascular atribuible a los remanentes de colesterol estimados con la fórmula de Martin y a la resistencia a la insulina (HOMA2-IR). Con base en los resultados, se puede decir que por cada aumento en una desviación estándar de remanentes de colesterol aumenta en un 28.96 % la incidencia relativa de muerte a causa de un evento cardiovascular. En cambio, un  $\text{HOMA2IR} \geq 2.5$  aumentó en

un 18.0% la incidencia relativa de muerte a causa de un evento cardiovascular en sujetos sin diabetes independientemente de la edad, sexo, IMC, etnicidad, LDL-C, tabaquismo y número de comorbilidades. Además, se probó la interacción en el modelo de  $HOMA2IR * Rem_S$ , pero no fue estadísticamente significativa,  $p\text{-valor} = 0.7238$ , lo que indica que no existe un efecto modificador entre las variables y, por tanto, es factible explorar un posible efecto mediador. Además, se evaluó el  $C - statistic$  ( $C = 0.847$ ) que indica un modelo que tiene una discriminación buena, además al comparar los estadísticos  $C$  del modelo 1 y 2, se observa el mismo valor, por lo tanto, discriminan de igual forma a sujetos que presentan el evento de interés. Como otra mediada de comparación entre modelos, se calculó el Criterio de Información Bayesiana ( $BIC$ ) y, prácticamente fue el mismo para ambos modelos; 19590.76 vs 19591.11, respectivamente.

Variable	$\hat{\beta}$	sHR	IC95 %	$p - value$
$Rem - M^+$	0.254	1.289	1.026-1.621	0.0293
$HOMA2IR \geq 2.5$	0.180	1.197	1.039-1.380	0.0130
Edad	0.068	1.070	1.066-1.075	<0.0001
Sexo (Mujer)	-0.248	0.779	0.692-0.879	<0.0001
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	-0.036	0.964	0.794-1.170	0.7131
Otro	-0.122	0.884	0.612-1.278	0.5132
Caucásico	-0.197	0.820	0.702-0.959	0.0130
IMC	0.012	1.012	0.999-1.025	0.0659
Comorbilidades	0.139	1.149	1.091-1.210	<0.0001
Tabaquismo	358	1.431	1.239-1.651	<0.0001
$LDL - C^+$	0.134	1.144	1.075-1.217	<0.0001
$Edad/Rem - M^*$	-0.004	0.992	0.992-0.999	0.0217
n=146458, eventos=1166	$C - statistic = 0.850$	$SE = 0.005$	BIC=19591.11	

Tabla 12: Modelo 2 de riesgos competitivos de Fine y Gray para evaluar mortalidad cardiovascular. Rem-M: Remanentes de colesterol-Martin. sHR: subdistribution Hazard ratio.\* Interacción; +los valores fueron centrados y estandarizados.

Posteriormente, se analizó a detalle la interacción entre remanentes de colesterol y la edad de los sujetos. En la *Figura 3.17*, se observa que en pacientes menores de 70 años a mayor concentración de remanentes de colesterol calculados se asocian con una mayor incidencia relativa de muerte a causa de un evento cardiovascular. En cambio, en sujetos con una edad igual o mayor a 70 años, una concentración mayor de remanentes de colesterol se asocia con una disminución discreta de la incidencia relativa de muerte a causa de un evento cardiovascular; aunque dichos sujetos tienen mayor riesgo de mortalidad cardiovascular en comparación con sujetos menores de 70 años.

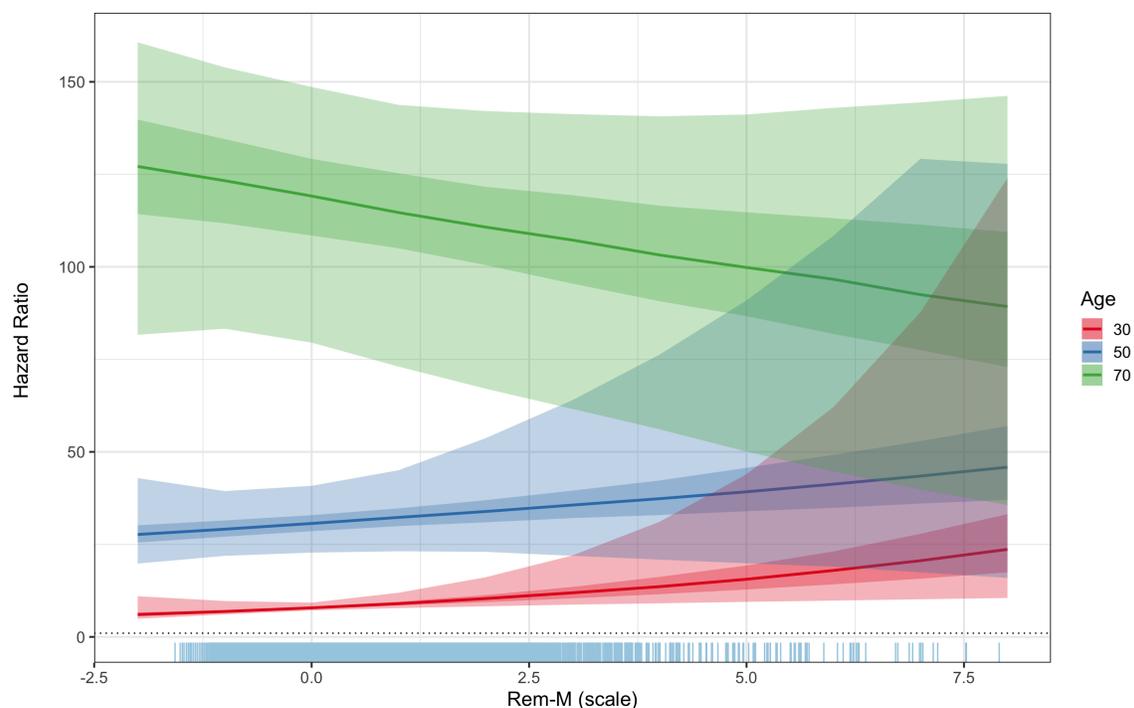


Figura 3.17: Evaluación de la interacción entre edad y remanentes de colesterol sobre la mortalidad cardiovascular en sujetos sin diabetes. Rangos de edad: 30 – 49, 50 – 69 y  $\geq 70$  años.

Para evaluar el ajuste del modelo se analizaron de forma visual los residuos de Schoenfeld, *Figura 3.18*. Como se puede observar, la distribución de los residuos para

cada variable independiente incluida en el modelo no muestra una tendencia clara. Sin embargo, al aplicar la función `cox.zph()`, se observa que para el caso de HOMA2IR y remanentes de colesterol, el  $p$ -valor es mayor a 0.05; 0.664 y 0.174, respectivamente. Lo que indica independencia de los residuales y el tiempo, pero el resto de variables y el análisis global tiene un  $p$ -valor  $\leq 0.0001$  lo que indica que hay una correlación entre el tiempo y los residuales.

Sin embargo, el tamaño de muestras y que el modelo de riesgos competitivos de Fine y Gray es una adaptación del modelo de riesgos proporcionales de Cox podría estar influyendo el resultado de la prueba, a pesar de que visualmente no se observa una tendencia clara. Se concluye a pesar de las limitantes, que el modelo ajusta adecuadamente los datos.

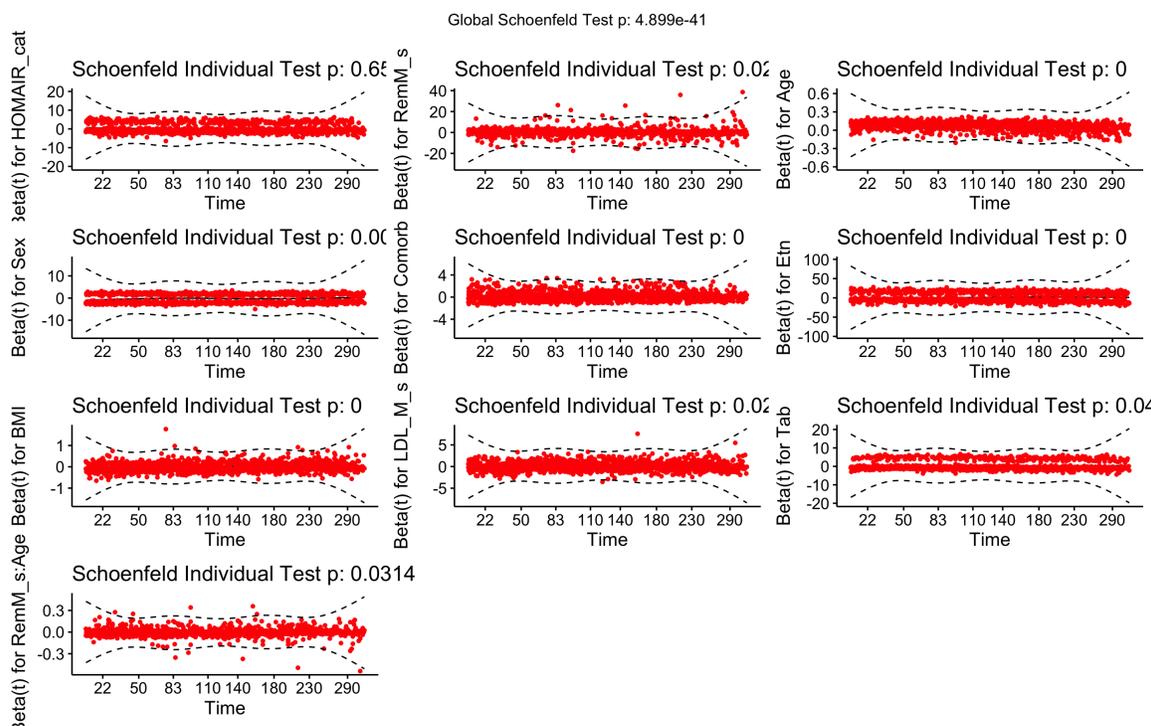


Figura 3.18: Supuesto de riesgos proporcionales: residuos de Schoenfeld.

Para evaluar posibles observaciones influyentes, se analizaron los valores  $dfbeta$  para cada coeficiente utilizando, *Figura 19*. Al analizar los gráficos se observa que no existen valores realmente influyentes, los coeficientes de cada variable cambian muy

poco al eliminar ciertas observaciones. Por lo tanto, no es necesario limitar aún más la muestra.

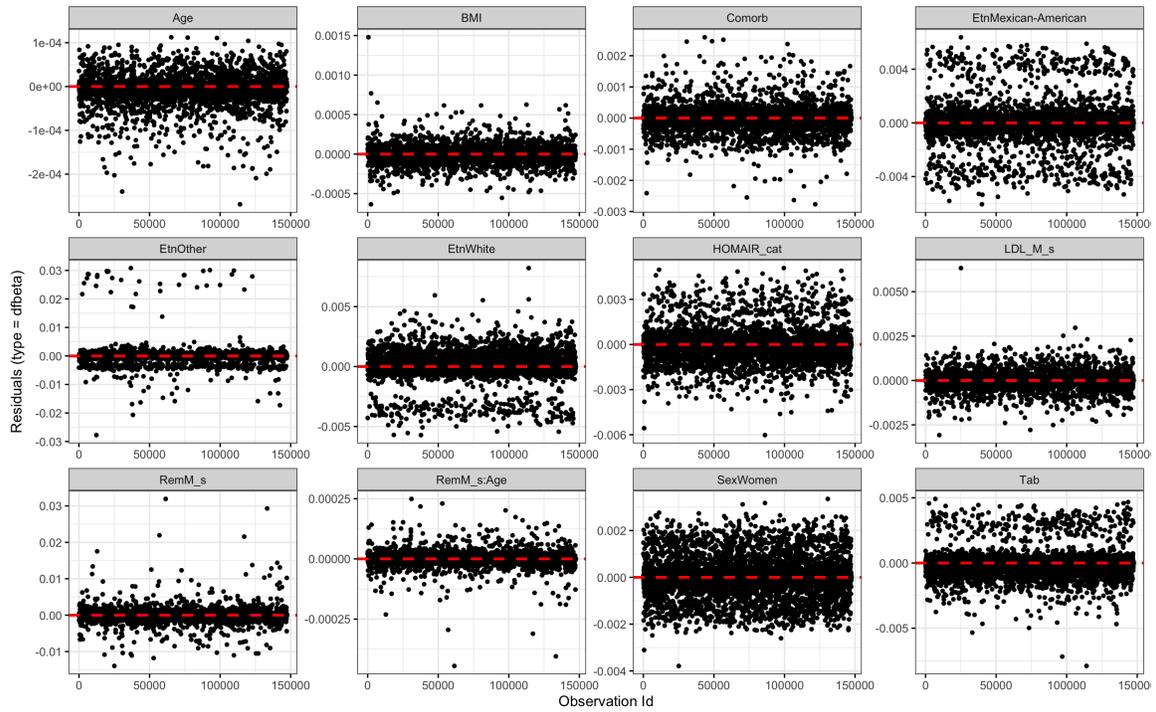


Figura 3.19: Gráficos de residuales (*dfbeta*) para evaluar posibles observaciones influyentes en el modelo de riesgos competitivos.

### 3.2. Modelos de riesgos competitivos de Fine y Gray en sujetos con diabetes

Es importante considerar que un factor relevante de la fisiopatología de la diabetes mellitus tipo 2 es la resistencia a la insulina. Por lo tanto, los sujetos con diabetes tendrán un HOMA2IR más alto en comparación con sujetos sin diabetes, pero en algunos casos dada la falla de la célula beta es esperado que se muestren valores bajos de HOMA2IR debido a la pérdida de la función de la célula beta que condicionará a una menor o nula síntesis y secreción de insulina y péptido C, parámetro utilizado para calcular HOMA2IR en esta muestra. Entonces, sujetos con una enfermedad en menor control o con un curso de la enfermedad más avanzado podrían tener un HOMA2IR bajo y tener mayor riesgo de mortalidad cardiovascular o por cualquier causa.

Ergo, es esperado que HOMA2IR sea un menor predictor de mortalidad cardiovascular en comparación con otros parámetros bioquímicos como la hemoglobina glucosilada. Por estas razones se decidió analizar por separado a los sujetos con diabetes. A continuación se presentan los modelos ajustado para este subgrupo de la muestra.

#### 3.2.1. Modelos 3 de riesgos competitivos de Fine y Gray en sujetos con diabetes

En primer lugar, se evaluó si la resistencia a la insulina evaluada mediante HOMA2IR y los remanentes de colesterol (Rem-S) conferían un aumento en la incidencia relativa de mortalidad cardiovascular. Un  $\text{HOMA2IR} \geq 4.9$  como el aumento en una desviación estándar de Rem-S se relacionó con un aumento significativo de la incidencia relativa de mortalidad causa-específica (cardiovascular); [sHR=1.379, IC95 % 1.056-1.800] y [sHR=1.340, IC95 % 1.066-1.682], respectivamente. Sin embargo, al ajustar por edad, sexo, IMC, etnicidad, LDL-C, tabaquismo y número de comorbilidades, *Tabla 13*, para HOMA2IR no se observó una asociación estadísticamente significativa con la mortalidad cardiovascular; [sHR=1.232, IC95 % 0.936-1.621]; en

cambio, tener una concentración  $\geq 40\text{mg/dL}$  de remanentes de colesterol aumentará en un 28.5% la incidencia relativa de mortalidad cardiovascular en pacientes con diabetes independientemente de la edad, sexo, IMC, comorbilidades, LDL-C, etnicidad y tabaquismo; [sHR=1.285, IC95% 1.016-1.626].

Además, se probó la interacción en el modelo de  $HOMA2IR * Rem_S$ , pero no fue estadísticamente significativa,  $p\text{-valor} = 0.9290$ , lo que indica que no existe un efecto modificador entre las variables y, por tanto, es factible explorar un posible efecto mediador. Al evaluar la interacción de remanentes de colesterol y edad, en sujetos con diabetes no se observó que la edad tuviera un efecto modificador sobre los remanentes de colesterol,  $p\text{-valor} = 0.559$ . Por último, se evaluó el  $C\text{-statistic}$  ( $C = 0.711$ ) que indica un modelo con una discriminación adecuada.

Para evaluar el ajuste del modelo se analizaron de forma visual los residuos de Schoenfeld, *Figura 3.20*. Como se puede observar, la distribución de los residuos para cada variable independiente incluida en el modelo no muestra una tendencia clara. Sin embargo, al aplicar la función `cox.zph()`, se observa que para el caso de HOMA2IR, remanentes de colesterol y sexo, el  $p\text{-valor}$  es mayor a 0.05; 0.170, 0.513 y 0.657, respectivamente. Lo que indica independencia de los residuales y el tiempo, pero el resto de variables y el análisis global tiene un  $p\text{-valor} \leq 0.05$  indicando una correlación entre el tiempo y los residuales del resto de las variables.

Sin embargo, como se ha comentado en la sección previa, el modelo de riesgos competitivos de Fine y Gray es una adaptación del modelo de riesgos proporcionales de Cox y dado esto es factible que no se cumpla con el supuesto al aplicar la prueba, aunque visualmente no se observa una tendencia. Se concluye a pesar de las limitantes, que el modelo ajusta adecuadamente los datos.

Variable	$\hat{\beta}$	sHR	IC95 %	$p - value$
$Rem - S \geq 40.0 \text{ mg/dL}$	0.251	1.285	1.016-1.626	0.0364
$HOMA2IR \geq 4.9$	0.208	1.232	0.936-1.621	0.1361
Edad	0.034	1.035	1.025-1.045	<0.0001
Sexo (Mujer)	-0.168	0.844	0.672-1.061	0.1474
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	0.016	1.016	0.733-1.409	0.9233
Otro	-0.053	0.947	0.453-1.980	0.8868
Caucásico	0.228	1.256	0.951-1.660	0.1076
IMC	-0.014	1.013	0.966-1.008	0.2143
Comorbilidades	0.202	1.224	1.130-1.326	<0.0001
Tabaquismo	-0.032	0.967	0.716-1.309	0.8354
LDL-C	0.097	1.103	0.986-1.234	0.0877
n=15985, eventos=329	$C - statistic = 0.711$	$SE = 0.013$	BIC=4651.155	

Tabla 13: Modelo 3 de riesgos competitivos de Fine y Gray para evaluar mortalidad cardiovascular. Rem-S: Remanentes de colesterol-Sampson. sHR: subdistribution Hazard ratio.

Para evaluar posibles observaciones influyentes, se analizaron los valores  $dfbeta$  para cada coeficiente utilizando, *Figura 3.21*. Al analizar los gráficos se observa que no existen valores influyentes, los coeficientes de cada variable cambian muy poco al eliminar ciertas observaciones. Por lo tanto, no es necesario eliminar a sujetos del modelo.

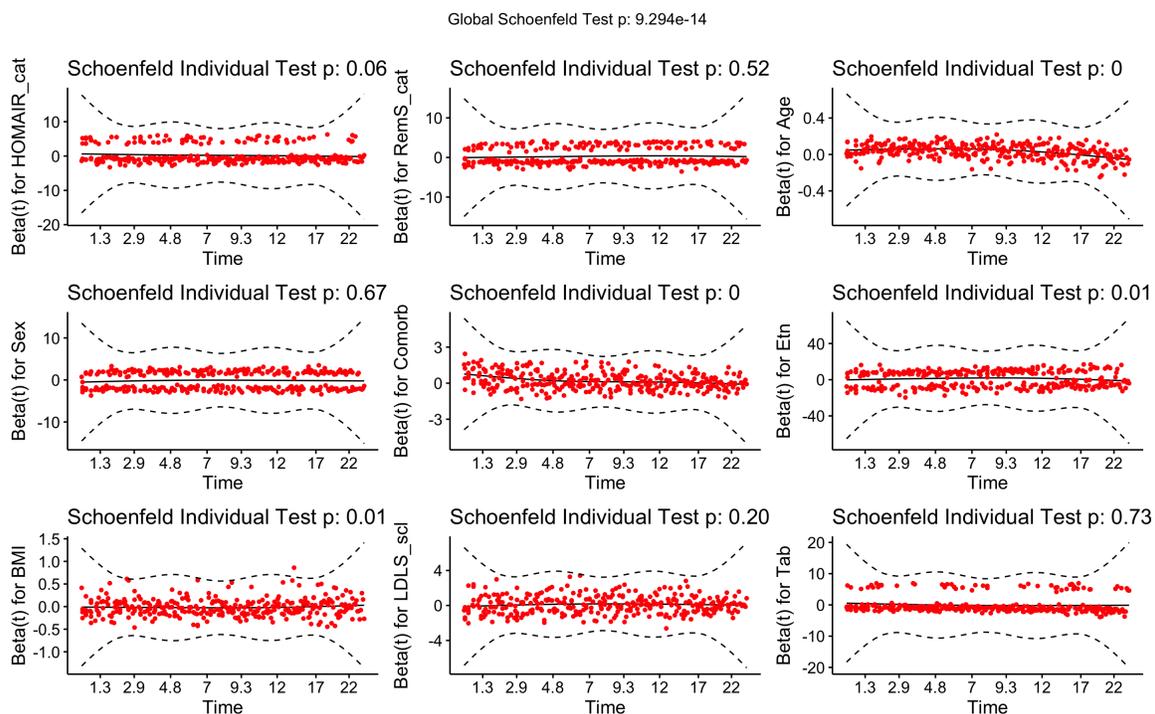


Figura 3.20: Supuesto de riesgos proporcionales: residuos de Schoenfeld.

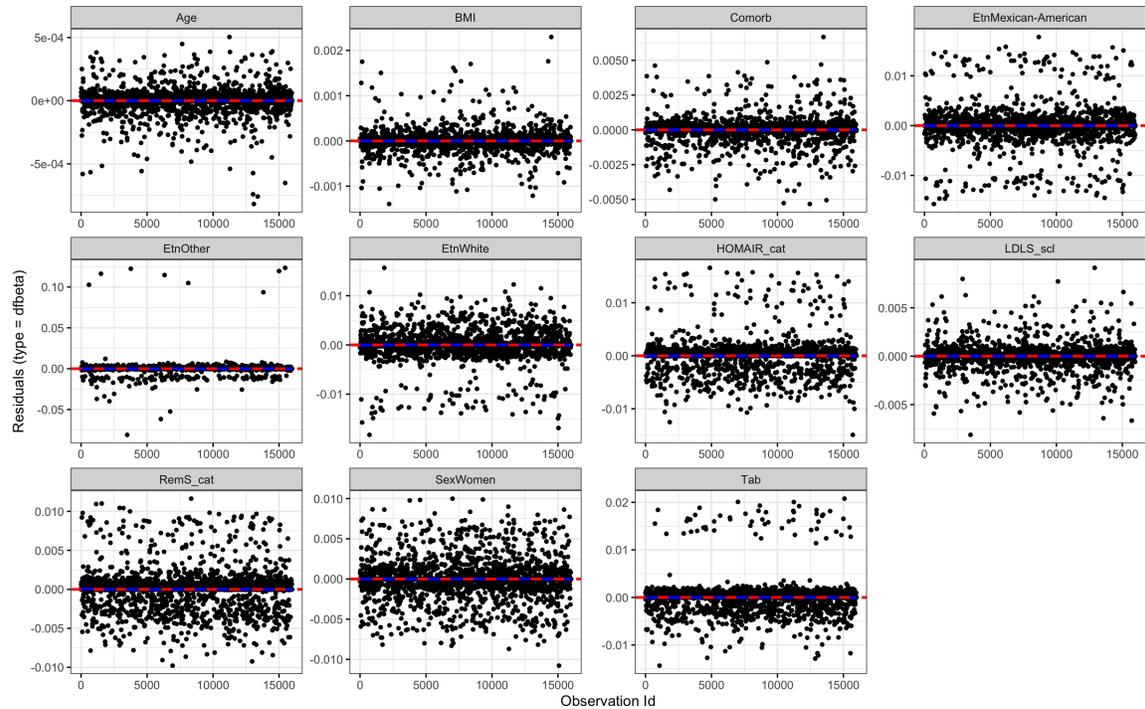


Figura 3.21: Gráficos de residuales (*dfbeta*) para evaluar posibles observaciones influyentes en el modelo de riesgos competitivos.

### 3.2.2. Modelo 4 de riesgos competitivos de Fine y Gray en sujetos con diabetes

De igual forma, se evaluó si la resistencia a la insulina (HOMA2IR) y los remanentes de colesterol (Rem-M) conferían un aumento en la incidencia relativa de mortalidad cardiovascular. Un  $HOMA2IR \geq 4.9$  como el aumento en  $Rem-M \geq 38 \text{ md/dL}$  se relacionaron con un aumento significativo de la incidencia relativa de mortalidad causa-específica (cardiovascular); [sHR=1.383, IC95 % 1.058-1.807] y [sHR=1.366, IC95 % 1.082-1.723], respectivamente. Sin embargo, al ajustar por edad, sexo, IMC, etnicidad, LDL-C, tabaquismo y número de comorbilidades, *Tabla 14*, para HOMA2IR no se observó una asociación estadísticamente significativa con la mortalidad cardiovascular; [sHR=1.217, IC95 % 0.923-1.606]; en cambio, tener una concentración  $\geq 40 \text{ mg/dL}$  de remanentes de colesterol aumentará en un 29.5 % la incidencia relativa de mortalidad cardiovascular en pacientes con diabetes independientemente de la edad, sexo, IMC, comorbilidades, LDL-C, etnicidad y tabaquismo; [sHR=1.295, IC95 % 1.011-1.659].

Además, se probó la interacción en el modelo de  $HOMA2IR * Rem_S$ , pero no fue estadísticamente significativa,  $p - \text{valor} = 0.6183$ , lo que indica que no existe un efecto modificador entre las variables y, por tanto, es factible explorar un posible efecto mediador. Al evaluar la interacción de remanentes de colesterol y edad, en sujetos con diabetes no se observó que la edad tuviera un efecto modificador sobre los remanentes de colesterol,  $p - \text{valor} = 0.7080$ . Por último, se evaluó el  $C - \text{statistic}$  ( $C = 0.711$ ) que indica un modelo con una discriminación adecuada.

Para evaluar el ajuste del modelo se analizaron de forma visual los residuos de Schoenfeld, *Figura 3.22*. Como se puede observar, la distribución de los residuos para cada variable explicativa incluida en el modelo no muestra una tendencia. Sin embargo, al aplicar la función `cox.zph()`, se observa que para el caso de HOMA2IR, remanentes de colesterol y sexo, el  $p - \text{valor}$  es mayor a 0.05; 0.173, 0.463 y 0.684,

respectivamente. Lo que indica independencia de los residuales y el tiempo, pero el resto de variables y el análisis global tiene un  $p - valor \leq 0.05$  lo que indica que hay una correlación entre el tiempo y los residuales del resto de las variables.

Variable	$\hat{\beta}$	sHR	IC95 %	$p - value$
$Rem - M \geq 38.0mg/dL$	0.258	1.295	1.011-1.659	0.0408
$HOMA2IR \geq 4.9$	0.197	1.217	0.923-1.606	0.1635
Edad	0.035	1.036	1.026-1.045	<0.0001
Sexo (Mujer)	-0.169	0.843	0.674-1.057	0.1399
Afroamericanos	Ref	Ref	Ref	-
Mexicoamericanos	0.007	1.007	0.726-1.396	0.9577
Otro	-0.036	0.964	0.464-2.002	0.9228
Caucásico	0.222	1.248	0.945-1.649	0.1173
IMC	-0.013	0.987	0.966-1.008	0.1399
Comorbilidades	0.201	1.223	1.129-1.325	<0.0001
Tabaquismo	-0.027	0.973	0.719-1.316	0.8597
LDL-C	0.097	1.083	0.966-1.215	172
n=16042, eventos=329	$C - statistic = 0.711$	$SE = 0.013$	BIC=4651.154	

Tabla 14: Modelo 4 de riesgos competitivos de Fine y Gray para evaluar mortalidad cardiovascular. Rem-M: Remanentes de colesterol-Martin. sHR: subdistribution Hazard ratio.

Sin embargo, como se ha comentado en la sección previa, el modelo de riesgos competitivos de Fine y Gray es una adaptación del modelo de riesgos proporcionales de Cox y dado esto es factible que no se cumpla con el supuesto al aplicar la prueba, aunque visualmente no se observa una tendencia. Se concluye a pesar de las limitan-

tes, que el modelo ajusta adecuadamente los datos.

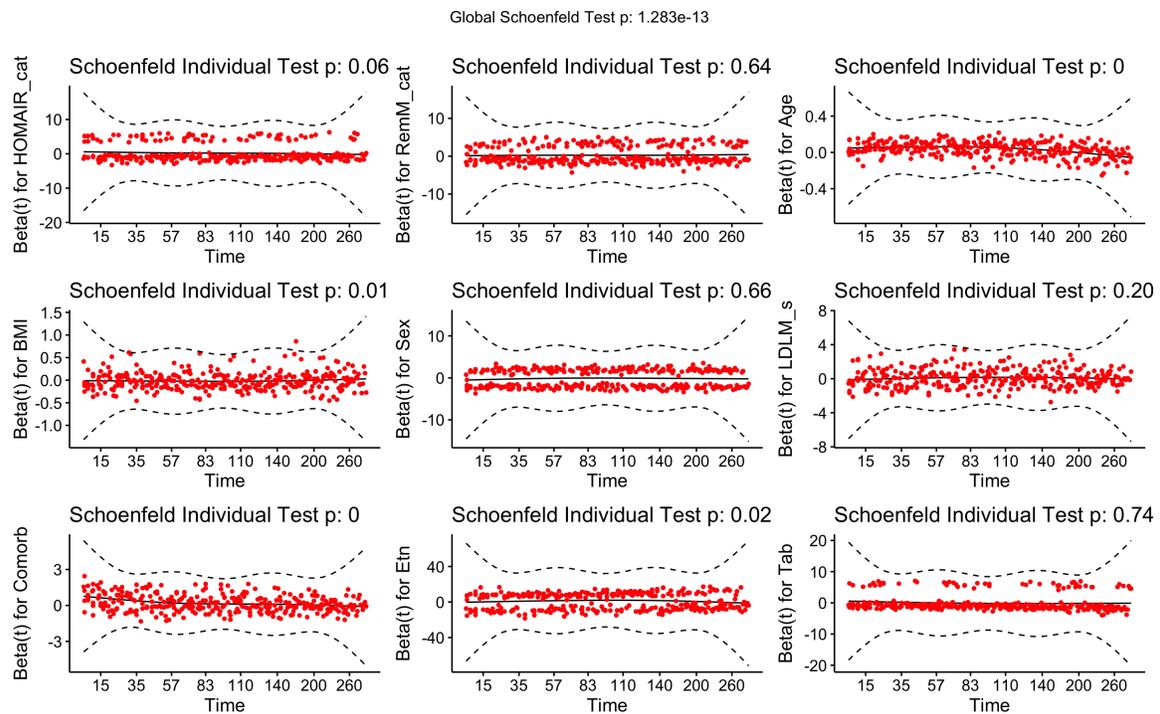


Figura 3.22: Supuesto de riesgos proporcionales: residuos de Schoenfeld.

Para evaluar posibles observaciones influyentes, se analizaron los valores  $dfbeta$  para cada coeficiente utilizando, *Figura 23*. Al analizar los gráficos se observa que no existen valores influyentes, los coeficientes de cada variable cambian muy poco al eliminar ciertas observaciones. Por lo tanto, no es necesario eliminar a sujetos del análisis.

Por último, se compararon los modelos 3 y 4 acorde al  $C - statistic$  y el BIC. En este caso, ningún modelo fue superior al otro dado que ambas estadísticas utilizadas son similares; (BIC; 4651.155 vs 4651.154) y ( $C$ -statistic; 0.711 vs 0.711).

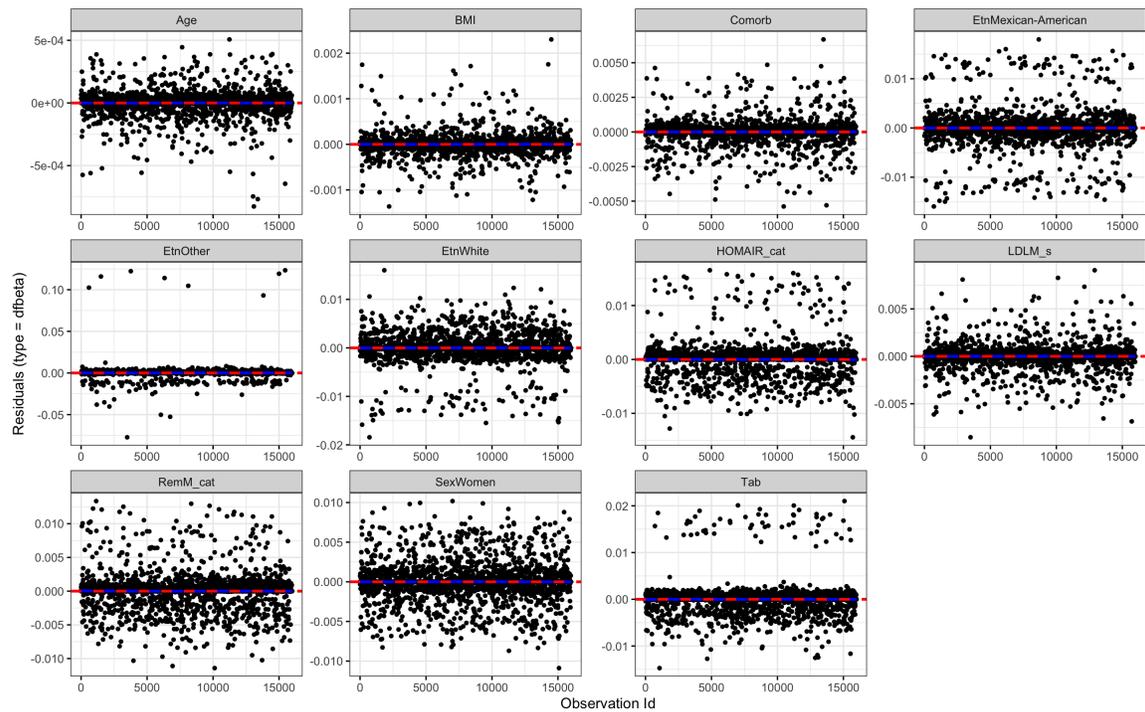


Figura 3.23: Gráficos de residuales (*dfbeta*) para evaluar posibles observaciones influyentes en el modelo de riesgos competitivos.

## 4. Análisis de Mediación Causal

Partiendo del modelo presentado en la *Figura 2.3*, donde se especifica que los remanentes de colesterol son un mediador de la relación entre la resistencia a la insulina y la mortalidad cardiovascular. Se ajustaron modelos de mediación causal partiendo de la suposición de que no existen factores de confusión no medidos para la relación exposición-evento, el mediador-evento y exposición-mediador. Además, se consideró que la exposición (HOMA2IR) tiene su efecto a través de 2 vías causales distintas; efecto directo y un efecto indirecto (a través del mediador), y no entrelazadas, por ende no existe otra variable que se vea afectada por la exposición y a su vez influya en el mediador (remanentes de colesterol). Así mismo, se asumió que los modelos de RLM y de riesgos competitivos de Fine y Gray ajustaron adecuadamente a los datos y que estos cumplían con los supuestos establecidos para cada modelo.

### 4.1. Análisis de Mediación Causal en sujetos sin diabetes

Del modelo ajustado en sujetos sin diabetes, el efecto total dado de la resistencia a la insulina sobre la mortalidad cardiovascular fue de  $sHR_S = 1.285$  y  $sHR_M = 1.295$ , por lo tanto, se puede concluir que un  $HOMA2IR \geq 2.5$  aumenta alrededor del 29% la incidencia relativa de mortalidad a causa de un evento cardiovascular. Sin embargo, con el modelo de mediación podemos discernir entre el efecto natural directo dado por la resistencia a la insulina sobre mortalidad cardiovascular ( $X \text{ --- } > Y$ ) y el efecto natural indirecto a través del mediador (remanentes de colesterol);  $X \text{ --- } > M \text{ --- } > Y$ .

Dado el modelo propuesto, la resistencia a la insulina se asocia con un 27.9-31.3% de aumento en la incidencia realtiva de mortalidad cardiovascular en sujetos sin diabetes (efecto directo natural). Sin embargo, la resistencia a la insulina ejerce un efecto a través de los remanentes de colesterol (mediador): 7.6-6.7% de aumento de la

incidencia relativa de mortalidad cardiovascular. Que corresponde a una proporción de mediación de alrededor de 18.36 %.

Modelo	Efecto total (sHR)	Efecto directo (sHR)	Efecto indirecto (sHR)	Proporción mediada
Rem-S	1.411 (1.120-1.778)	1.313 (1.046-1.644)	1.076 (1.023-1.131)	18.36 (16.84-19.28)
Rem-M	1.366 (1.080-1.725)	1.279 (1.016-1.610)	1.067 (1.021-1.115)	18.33 (15.95-25.86)

Tabla 15: Análisis de mediación causal en sujetos sin diabetes. Los modelos están ajustados por edad, sexo, IMC, etnicidad y número de comorbilidades. Rem-M: Remanentes de colesterol-Martin; Rem-S: Remanentes de colesterol-Sampson. Los sHR se expresan junto con su respectivo IC95 % .

En resumen, del efecto total (100 %) dado por la resistencia a la insulina sobre mortalidad cardiovascular, un 18 % se debe al efecto indirecto ejercido a través de la relación *resistencia a la insulina* – – – > *remanentes de colesterol* – – – > *mortalidad cardiovascular*, *Tabla 15*.

## 4.2. Análisis de Mediación Causal en sujetos con diabetes

Para el caso de la muestra de pacientes con diabetes, se probó un modelo de mediación para el modelo simple (HOMA2IR y Remanentes de colesterol) y un segundo modelo que considera las variables de ajuste (modelo saturado). Para el caso del modelo simple, se observó que el efecto total de la resistencia a la insulina sobre la mortalidad cardiovascular fue de  $RR_S = 1.493$  y  $RR_M = 1.514$ , es decir, la resistencia a la insulina aumenta en alrededor de 50 % la incidencia relativa de mortalidad cardiovascular en sujetos con diabetes. Sin embargo, la resistencia a la insulina ejerce un efecto a través de los remanentes de colesterol (mediador): 10.9-11.5 % de aumento de la incidencia relativa de mortalidad cardiovascular. Que corresponde a una proporción de mediación de alrededor de 20 %, *Tabla 16*.

Modelo	Efecto total	Efecto directo	Efecto indirecto	Proporción mediada
Rem-S	1.493 (1.186-1.879)	1.338 (1.065-1.681)	1.115 (1.018-1.225)	23.40 (9.98-25.57)
Rem-M	1.514 (1.200-1.909)	1.365 (1.082-1.772)	1.109 (1.018-1.211)	21.20 (9.02-23.18)

Tabla 16: Análisis de mediación causal en sujetos con diabetes. Modelos simples. Rem-M: Remanentes de colesterol-Martin; Rem-S: Remanentes de colesterol-Sampson.

Aunque, al considerar el modelo saturado. En sujetos con diabetes, el efecto total de la resistencia a la insulina sobre la mortalidad cardiovascular fue  $sHR_S = 1.379$  y  $sHR_M = 1.377$ ; por lo tanto, se puede concluir que un  $HOMA2IR \geq 4.9$  aumenta alrededor del 38% al incidencia relativa de morir a causa de un evento cardiovascular. Al considerar el modelo de mediación se tiene que el efecto natural directo dado por la resistencia a la insulina sobre mortalidad cardiovascular ( $X \text{ --- } > Y$ ) es  $sHR_S = 1.285$  y  $sHR_M = 1.294$ . En cambio, el efecto natural indirecto a través del mediador (remanentes de colesterol);  $X \text{ --- } > M \text{ --- } > Y$  está entre  $sHR_S = 1.074$  y  $RR_M = 1.064$ , *Tabla 17*. Sin embargo, el efecto indirecto no es estadísticamente significativo, por lo tanto, en pacientes con diabetes la resistencia a la insulina tiene un efecto directo total sobre mortalidad cardiovascular y los remanentes de colesterol no median dicha relación.

Modelo	Efecto total	Efecto directo	Efecto indirecto	Proporción mediada
Rem-S	1.379 (1.090-1.745)	1.285 (1.015-1.625)	1.074 (0.978-1.180)	21.9.00 (4.95-24.22)
Rem-M	1.377 (1.082-1.753)	1.294 (1.010-1.656)	1.064 (0.974-1.164)	16.89 (1.32-30.92)

Tabla 17: Análisis de mediación causal en sujetos con diabetes. Los modelos ajustados por edad, sexo, IMC, etnicidad, tabaquismo, LDL-C y número de comorbilidades. Rem-M: Remanentes de colesterol-Martin; Rem-S: Remanentes de colesterol-Sampson.

# Capítulo 4

## Discusión y Conclusiones

Con base en los análisis presentados, se mostró que el efecto (total) que tiene la resistencia a la insulina sobre la mortalidad cardiovascular está mediado por los remanentes de colesterol en sujetos sin diabetes, no así en sujetos con diagnóstico de diabetes. Es decir, la resistencia a la insulina a través de ciertas vías metabólicas aumenta la producción de partículas *VLDL*, así como un menor aclaramiento de las mismas y de los remanentes de quilomicrones, que a su vez una mayor concentración de remanentes de colesterol (*VLDL-C*, *IDL-C* y *remanentes de quilomicrones*) condicionan un mayor depósito sobre las paredes arteriales y, por ende, contribuyen a la formación de la placa aterosclerótica. Además, los remanentes de colesterol afectan la reactividad de los vasos sanguíneos. Por lo tanto, un mayor riesgo de mortalidad cardiovascular [4, 8].

El uso de los modelos de mediación causal en investigación clínica permite dilucidar la relación entre un evento/efecto de interés y las posibles variables explicativas, que en ocasiones puede resultar en una relación compleja de entender usando métodos tradicionales. Por ejemplo, los modelos de RLM por sí mismos no logran discernir los mecanismos subyacentes a la asociación entre la variable explicativa y el evento/resultado de interés, mientras que los investigadores pueden estar interesados en tales procesos fisiopatológicos mediadores subyacentes [57]. Sin embargo, es necesario que el investigador tenga cierto grado de conocimiento previo de que la variable explicati-

va ejerce su efecto sobre el resultado a través de vías directas e indirectas. En el caso de la vía indirecta, un mediador que transmite el efecto causal [56, 57].

El uso de los modelos de mediación causal no sólo se limita al contexto de estudios observacionales, también es posible aplicarlo a estudios clínicos aleatorizados (ECA). En este contexto, los ECA permiten controlar ciertas condiciones (variables) bajo estudio, además es posible establecer relaciones causa-efecto si el diseño es adecuado. El uso del análisis de mediación permite confirmar el mecanismo de trabajo hipotético subyacente a una intervención, o examinar por qué una intervención no tuvo éxito al ayudar a identificar posibles mejoras para el diseño y la evaluación de futuros estudios de intervención [71].

Por otro lado, el uso de modelos de supervivencia dentro del análisis de mediación resalta la importancia de evaluar cómo las variables explicativas ejercen un efecto sobre cierto evento a lo largo del tiempo. En particular, el análisis de supervivencia permite estudiar mortalidad (fallo) y el tiempo de supervivencia con la consideración adecuada de los tiempos de seguimiento, la censura y los riesgos competitivos. La regresión de riesgos proporcionales de Cox es la técnica más utilizada, pero su uso en el análisis de mediación plantea algunos desafíos. Dicho modelo se basa en la proporcionalidad del riesgo y esta podría no cumplirse si se agrega una variable (medidora) a un modelo correctamente especificado. Por ende, la inclusión de la variable podría desplazar los riesgos de la línea de base, en lugar de sólo modificar la pendiente de la función de riesgo, fenómeno de la *no colapsabilidad* de la razón de riesgo [72, 73]. Por esta razón, el uso del modelo de Cox para estimar los efectos indirectos a través de la diferencia o el producto de los coeficientes se basa en el supuesto de que el resultado es raro (<10%) [59, 84].

Al considerar que la ocurrencia de eventos distintos al de interés que condicionan la falla durante el seguimiento impactan sobre la estimación del riesgo atribuible a una variable explicativa, se ajustaron modelos de Fine y Gray que son una adapta-

ción del modelo de Cox para tener en cuenta la presencia de riesgos competitivos y mantener las observaciones en el conjunto de riesgo con una ponderación decreciente. Por lo tanto, proporciona un modelo de los riesgos de subdistribución. Sin embargo, el supuesto de proporcionalidad de los riesgos continúa siendo una exigencia, pero se refiere a los riesgos de subdistribución [40]. Al ajustar modelos de riesgos competitivos se observó que tanto la resistencia a la insulina como una mayor concentración de remanentes de colesterol estimados con las fórmulas de Sampson y Martin se asociaron con un mayor riesgo de mortalidad cardiovascular comparado con otras posibles causas de mortalidad.

Diversos estudios muestran que el contenido de colesterol en las lipoproteínas ricas en triglicéridos y sus remanentes se asocian con riesgo y/o mortalidad cardiovascular incluso en pacientes con niveles óptimos de colesterol LDL e independientemente de los niveles de apolipoproteína B-100 (proteína estructural de los lipoproteínas)[85]. Ajustando modelos de riesgos competitivos mostramos que el contenido de colesterol en los remanentes, así como la resistencia a la insulina evaluada por HOMA2IR se asocian con mayor riesgo de mortalidad tanto en sujetos con o sin diabetes, independientemente de la edad, sexo, IMC, número de comorbilidades, tabaquismo, colesterol LDL y etnicidad. Lo que apoya a la creciente evidencia que sugiere que los triglicéridos (TG) y/o el contenido de colesterol dentro de las lipoproteínas ricas en triglicéridos (TRL) pueden contribuir al riesgo residual de enfermedad cardiovascular aterosclerótica. A pesar de que la relación en TGs y enfermedad cardiovascular aterosclerótica es controversial. La presencia de hipertrigliceridemia se asocia con alteraciones en la composición de las lipoproteínas (su contenido de colesterol y triglicéridos), además con niveles bajos de colesterol HDL y niveles elevados de colesterol no HDL; que incluye partículas de LDL pequeñas, densas y totales; lipoproteína de densidad intermedia (IDL); y apolipoproteína C-III. Las cuales están asociados con un mayor riesgo de cardiovascular [86, 87].

En 2007 Kato T, et al. evaluaron el efecto de la concentración sérica de remanen-

tes de colesterol y la resistencia a la insulina en 64 pacientes sin diabetes con angina estable que fueron sometidos a un procedimiento de colocación de stent coronario. Después de un promedio de 6 meses de seguimiento observaron re-estenosis intra-stent en el 23% de los sujetos, interesantemente los sujetos con re-estenosis tuvieron una concentración más alta de remanentes de colesterol, así como un HOMAIR significativamente más alto. Sin embargo, en el modelo de regresión logística multivariado, sólo los remanentes de colesterol fueron un factor de riesgo independiente asociado a la re-estenosis [74]. A pesar de ser uno de los primeros estudios que evaluó de forma conjunta los remanentes de colesterol y la resistencia a la insulina. El estudio tiene ciertas limitantes que deben considerarse, primero el tamaño de muestra pequeño ( $n = 64$ ) pudo limitar la descripción del fenómeno, incluir dentro del modelo de regresión logística tanto insulina en ayuno como HOMA2IR supone una alta probabilidad de colinealidad en el modelo, dado que el cálculo de HOMAIR incluye a la insulina. Además, no se evaluó una posible interacción entre HOMAIR y remanentes de colesterol ni un posible efecto mediador, aunque el ajuste de dichos modelos pudo estar limitado por el tamaño de muestra. En cambio, los análisis presentados en este trabajo muestran una relación lineal positiva entre HOMA2IR y remanentes de colesterol y, además que tanto HOMA2IR como los remanentes de colesterol son factores de riesgo asociados a mortalidad cardiovascular tanto en sujetos con y sin diabetes independientemente de la edad, sexo, IMC, etnicidad, tabaquismo, colesterol LDL y número de comorbilidades.

En cambio, Goudriaan et al. describió que el receptor VLDL, que en realidad es el receptor para los remanentes de VLDL más importante, tiene un papel clave en la obesidad y la resistencia a la insulina en ratones alimentados con una dieta alta en grasas y azúcar refinada (HFS). Se ha descrito previamente que una HFS es capaz de inducir obesidad y resistencia a la insulina, así como hiperlipidemia. Interesantemente, en ratones knockout para el receptor de VLDL que reciben una HSF se observa un aumento de TG (remanentes de VLDL) en plasma, sin embargo no son capaces de almacenar TG en el tejido adiposo y, a pesar de la HSF no desarrollan resistencia

a la insulina. Por lo tanto, la formación de remanentes de VLDL como primer paso después de una dieta HSF debe colocarse antes de la obesidad y la resistencia a la insulina o al menos en la misma posición que la glucosa en la sangre [75, 76, 77]. Dado esto, es importante explorar un modelo que explore si la resistencia a la insulina es un mediador del efecto de los remanentes de colesterol sobre mortalidad cardiovascular.

Interesantemente, algunos estudios han analizado la relación de los niveles de colesterol y riesgo de mortalidad en sujetos mayores a 65 años de edad y han reportado que aquellos con niveles más altos de colesterol total tienen menor riesgo de mortalidad por cualquier causa [78]. Por otro lado, en el estudio de Rotterdam se encontró que un colesterol total más alto se asoció con un menor riesgo de mortalidad no-cardiovascular en adultos mayores y la fuerza de la asociación inversa aumentó con cada década de aumento en la edad [79]. Es decir, la reducción de mortalidad por cualquier causa se debe a una reducción de la mortalidad no-cardiovascular y no necesariamente a la reducción de mortalidad cardiovascular [78]. En contraste, los niveles de colesterol LDL se asocian a un mayor riesgo de infarto al miocardio y enfermedad cardiovascular incluso en sujetos mayores de 70 años de edad, además el número necesario a tratar para prevenir un evento cardiovascular en este rango de edad es menor en comparación con edades más tempranas. Sin embargo, el riesgo de enfermedad cardiovascular aterosclerótica sigue siendo mayor a menor edad, sujetos con mayor concentración de colesterol LDL entre 20-49 años de edad tienen mayor riesgo de eventos cardiovasculares y de mortalidad en comparación con aquellos mayores de 70 años [80]. En cambio, en la muestra analizada se observó una interacción estadísticamente significativa entre edad y remanentes de colesterol y al analizar dicha relación se observó que en sujetos mayores de 70 años de edad y sin diabetes el riesgo de subdistribución (sHR) disminuye a medida que aumentan los niveles de remanentes de colesterol no así para sujetos menores de 70 años, en quienes se observó que una mayor concentración se asoció con un aumento lineal de la incidencia relativa de mortalidad cardiovascular. Sin embargo, en sujetos con diabetes no se observó dicha interacción con la edad.

Por otro lado, en sujetos con diabetes los remanentes de colesterol no son un mediador de la relación entre resistencia a la insulina y mortalidad cardiovascular. La resistencia a la insulina ejerce un efecto natural directo completo sobre mortalidad cardiovascular. A pesar de que en los modelos saturados, la resistencia a la insulina no mostró una asociación significativa con un aumento de la incidencia relativa de mortalidad cardiovascular, diversos estudios han reportado que HOMAIR un índice de resistencia a la insulina se asocia con mayor riesgo cardiovascular en pacientes con diabetes [81]; por lo tanto, no es posible descartar dicha relación dado que no se observó una relación significativa en los modelos ajustados previamente. Además, el colesterol LDL no mostró una asociación estadísticamente significativa con una mayor incidencia relativa de mortalidad cardiovascular; en cambio, los remanentes de colesterol aumentaron el riesgo de subdistribución de mortalidad cardiovascular en pacientes con diabetes. Debido a que HOMAIR utiliza insulina o péptido C para la estimación de resistencia a la insulina; en sujetos con diabetes, IMC bajo y disfunción de célula  $\beta$  tiene un papel limitado para identificar a sujetos con resistencia a la insulina[82]. Además un estudio en población asiática reportó que el índice triglicéridos/glucosa es un mejor indicador de resistencia a la insulina en comparación con HOMAIR al comparar con el estándar de oro, clamp euglicémico hiperinsulinémico [83]. Esto insta a interpretar con precaución el efecto reportado de la resistencia a la insulina evaluado por HOMA2IR sobre la mortalidad cardiovascular en pacientes con diabetes y promueve la evaluación de la relación del descontrol metabólico (HbA1C) como un mejor marcador de mortalidad en estos sujetos e incluso explorar un posible efecto mediador de los remanentes de colesterol sobre la relación descontrol hiperglucémico y mortalidad cardiovascular.

A pesar de que diversos autores han aplicado modelos de mediación causal para entender la relación que tienen ciertas variables sobre el riesgo cardiovascular en muestras con un tamaño importante, así como con un tiempo de seguimiento adecuado. El uso actual de estos métodos es limitado debido a la pobre confianza de los análisis, debido principalmente a la falta de información sobre el cumplimiento de los supuestos de los modelos [84]. Debido a que el uso de análisis de mediación causal

en el análisis de supervivencia ha cobrado popularidad en investigación clínica y que el uso rutinario de métodos tradicionales ampliamente aceptados en la comunidad médica (por ejemplo el modelo de riesgos proporcionales de Cox) muchas veces no ajustan adecuadamente los datos, es necesario explorar otras técnicas de análisis de supervivencia que permitan mejorar la estimación del fenómeno estudiado. Además, es de suma importancia el desarrollo por parte de investigadores y expertos en metodología y estadística de un conjunto de criterios que permitan evaluar acorde al tipo de evento a estudiar, así como el método utilizado, la validez de los modelos de mediación causal.

### **Fortalezas y limitaciones**

Este trabajo tiene fortalezas y limitantes que deben considerarse. Primero, se presenta una revisión bibliográfica sobre los modelos utilizados en el análisis, así como la evaluación de los modelos a través del cumplimiento de los supuestos que implica cada uno. Cabe resaltar que se utilizó una muestra con gran número de sujetos con un tiempo de seguimiento importante, alrededor de 23 años (272(IQR 172 – 301) meses de seguimiento). Además, dada la información disponible en NHANES-III se pudo recuperar información basal de los sujetos así como mortalidad por cualquier causas y causa-específica, lo que permitió identificar casos de mortalidad cardiovascular. Sin embargo, no fue posible recuperar información sobre características clínicas, bioquímicas y/o antropométricas durante el seguimiento; por lo tanto, la estimación del riesgo se basa en una sólo medición (basal). Además, no se contó con la medición directa (estándar de oro) de remanentes de colesterol. Sin embargo, se calcularon los niveles de remanentes de colesterol a través de 2 fórmulas; Sampson y Martin, que tiene un desempeño adecuado en cuanto a la estimación de colesterol LDL y remanentes de colesterol en sujetos con y sin hipertrigliceridemia, principal factor que afecta la estimación de estos parámetros.

Debido al tamaño de muestra, fue posible segmentar a la muestra en dos grupos; sujetos con diabetes y sin diabetes. Por lo tanto, fue posible evaluar el fenómeno

de forma independiente del diagnóstico de diabetes. Aunque no fue posible incluir algunas variables que impactan en el nivel de remanentes de colesterol como el tratamiento con fibratos y/o estatinas, e incluso medicamentos hipoglucemiantes en sujetos con diabetes, lo que pudo limitar el desempeño de los modelos para explicar la variabilidad de los fenómenos estudiados. Es importante resaltar que debido al diseño de NHANES fue posible considerar como variable de ajuste la etnicidad en todos los modelos.

Durante la implementación de los modelos de RLM en sujetos sin diabetes, submuestra con mayor tamaño ( $n=10,990$ ), las pruebas estadísticas para evaluar el supuesto de homoscedasticidad y normalidad de los errores arrojaron un  $p$ -valor menor a 0.05 y, que por tanto, implicaba el rechazo de la hipótesis nula y la violación de dichos supuestos. Sin embargo, se optó por apegarse a la evaluación de la representación gráfica para ambos casos y concluir que dichos supuestos se cumplían. En un análisis de supervivencia tradicional, los eventos de mortalidad por otras causas se consideran como datos censurados y esto impacta en la estimación del riesgo atribuible a las covariables, debido a esto, se optó por implementar un modelo de riesgos competitivos de Fine y Gray, que al ser un modelo de riesgos proporcionales de Cox adaptado a los riesgos competitivos fue necesario evaluar el supuesto de riesgos proporcionales. A pesar de que al utilizar una prueba estadística para evaluar la correlación entre riesgo con el tiempo para cada variable incluida y para el modelo general, dicho supuesto no se cumplía. Se procedió a evaluar a través de la inspección visual la representación gráfica de los residuos de Schoenfeld y se concluyó que no había una tendencia con respecto a la relación del riesgo y el tiempo. Sin embargo, es importante considerarlo como una limitante del análisis que hace necesario explorar en futuros análisis el fenómeno a través de otras técnicas de análisis de supervivencia.

Cuando se usa el modelo de riesgos proporcionales de Cox, aunque en este caso un modelo de riesgos competitivos de Fine y Gray (adaptación del modelo de Cox), como parte del análisis de mediación. La estimación de los efectos directos e indirectos sólo

es adecuada cuando el evento de interés es poco frecuente ( $<10\%$ ). En este caso, la proporción de sujetos que murieron a causa de un evento cardiocascular en el caso de la submuestra de sujetos sin diabetes fue de  $10.7\%$ , en cambio para el caso sujetos con diabetes fue del  $25.9\%$ , lo que implicaría que el modelo no es aplicable para en el caso de sujetos con diabetes o incluso en ambos escenarios. Ergo, el tiempo de seguimiento fue de casi 23 años, por lo que al calcular las tasas de incidencia, se observó que para el caso de sujetos sin diabetes esta fue de aproximadamente 2 eventos por cada 1000 personas-año y para la submuestra de sujetos con diabetes de 6 eventos por cada 1000 personas-año. Entonces, se asumió que la frecuencia del evento de interés (mortalidad cardiovascular) fue baja y, además se considero un análisis de riesgos competitivos que proporciona un modelo de los riesgos de sub-distribución, por lo tanto, se asumió que el modelo de mediación era válido para estimar el efecto directo e indirecto, así como la proporción de mediación.

## Perspectivas

La aplicación de modelos de mediación causal en el contexto del análisis de supervivencia es una técnica útil cuando además de infomar el *por qué*, se busca el *cómo* de los resultados. Sin embargo, elegir el modelo de análisis de supervivencia resulta primordial para poder dar una interpretación al modelo de mediación o incluso asumir que éste es válido para estimar los efectos indirecto y directo, y por ende la proporción de mediación. En consecuencia, es necesario tener conocimiento de otras técnicas de análisis de supervivencia cuando el modelo de riesgos propocionales de Cox, modelo más utilizado en investigación clínica, no es válido debido a la falta de cumplimiento de los supuestos; o en su defecto reconcer las limitantes de los modelos cuando no se puede asegurar el cumplimiento adecuado de los supuestos. Todo ello, resalta la importancia de estandarizar un conjunto de criterios que permitan evaluar acorde al tipo de evento a estudiar, así como el método utilizado, la validez de los modelos de mediación causal y de esta forma aumentar la confianza en los resultados obtenidos a través de esta técnica de análisis. Por último, a pesar del ajuste de los modelos de mediación en NHANES-III, es necesario ajustar el modelo de mediación

causal planteado en otras poblaciones e incluso en otros estudios con una metodología distinta para replicar los resultados observados.

# Bibliografía

- [1] Faridi KF, Quispe R, Martin SS, Hendrani AD, Joshi PH, Brinton EA, Cruz DE, Banach M, Toth PP, Kulkarni K, Jones SR. Comparing different assessments of remnant lipoprotein cholesterol: The very large database of lipids. *J Clin Lipidol*. 2019 Jul-Aug;13(4):634-644. doi: 10.1016/j.jacl.2019.06.001.
- [2] Dash S, Xiao C, Morgantini C, Lewis GF. New Insights into the Regulation of Chylomicron Production. *Annu Rev Nutr*. 2015;35:265-94. doi: 10.1146/annurev-nutr-071714-034338.
- [3] Mittendorfer B, Yoshino M, Patterson BW, Klein S. VLDL Triglyceride Kinetics in Lean, Overweight, and Obese Men and Women. *J Clin Endocrinol Metab*. 2016 Nov;101(11):4151-4160. doi: 10.1210/jc.2016-1500.
- [4] Ginsberg HN, Packard CJ, Chapman MJ, Borén J, Aguilar-Salinas CA, Averna M, Ference BA, Gaudet D, Hegele RA, Kersten S, Lewis GF, Lichtenstein AH, Moulin P, Nordestgaard BG, Remaley AT, Staels B, Stroes ESG, Taskinen MR, Tokgözoğlu LS, Tybjaerg-Hansen A, Stock JK, Catapano AL. Triglyceride-rich lipoproteins and their remnants: metabolic insights, role in atherosclerotic cardiovascular disease, and emerging therapeutic strategies-a consensus statement from the European Atherosclerosis Society. *Eur Heart J*. 2021 Dec 14;42(47):4791-4806. doi: 10.1093/eurheartj/ehab551.
- [5] Howard BV. Insulin resistance and lipid metabolism. *Am J Cardiol*. 1999 Jul 8;84(1A):28J-32J. doi: 10.1016/s0002-9149(99)00355-0.

- [6] Wang Y, Viscarra J, Kim SJ, Sul HS. Transcriptional regulation of hepatic lipogenesis. *Nat Rev Mol Cell Biol.* 2015 Nov;16(11):678-89. doi: 10.1038/nrm4074. Erratum in: *Nat Rev Mol Cell Biol.* 2016 Jan;17(1):64.
- [7] Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, Chapman MJ, De Backer GG, Delgado V, Ference BA, Graham IM, Halliday A, Landmesser U, Mihaylova B, Pedersen TR, Riccardi G, Richter DJ, Sabatine MS, Taskinen MR, Tokgozoglu L, Wiklund O; ESC Scientific Document Group. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J* 2020;41:111–188.
- [8] Varbo A, Benn M, Nordestgaard BG. Remnant cholesterol as a cause of ischemic heart disease: evidence, definition, measurement, atherogenicity, high risk patients, and present and future treatment. *Pharmacol Ther.* 2014 Mar;141(3):358-67. doi: 10.1016/j.pharmthera.2013.11.008.
- [9] Nordestgaard, B. G., Wootton, R., Lewis, B. (1995). Selective retention of VLDL, IDL, and LDL in the arterial intima of genetically hyperlipidemic rabbits in vivo. Molecular size as a determinant of fractional loss from the intima-inner media. *Arterioscler Thromb Vasc Biol* 15, 534–542.
- [10] Nordestgaard, B. G., Stender, S., Kjeldsen, K. (1988). Reduced atherogenesis in cholesterol-fed diabetic rabbits. Giant lipoproteins do not enter the arterial wall. *Arteriosclerosis* 8, 421–428.
- [11] Brunzell, J.D., Deeb, S. S. (2001). Familial lipoprotein lipase deficiency, Apo C-II deficiency, and hepatic lipase deficiency. In C. R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle (Eds.), *The metabolic and molecular bases of inherited diseases* (pp. 2789–2816) (8th ed.). Mc-Graw-Hill, Medical Publishing Division.
- [12] Nakajima, K., Nakano, T., Tanaka, A. (2006). The oxidative modification hypothesis of atherosclerosis: The comparison of atherogenic effects on oxidized LDL and remnant lipoproteins in plasma. *Clin Chim Acta* 367, 36–47.

- [13] Zheng, X. Y., Liu, L. (2007). Remnant-like lipoprotein particles impair endothelial function: Direct and indirect effects on nitric oxide synthase. *J Lipid Res* 48, 1673–1680.
- [14] Wang, L., Gill, R., Pedersen, T. L., Higgins, L. J., Newman, J. W., Rutledge, J. C. (2009). Triglyceride-rich lipoprotein lipolysis releases neutral and oxidized FFAs that induce endothelial cell inflammation. *J Lipid Res* 50, 204–213.
- [15] Grant, P. J. (2007). Diabetes mellitus as a prothrombotic condition. *J Intern Med* 262, 157–172.
- [16] Varbo, A., Benn, M., Tybjaerg-Hansen, A., Nordestgaard, B. G. (2013). Elevated remnant cholesterol causes both low-grade inflammation and ischemic heart disease, while elevated low-density lipoprotein cholesterol causes ischemic heart disease without inflammation. *Circulation* 128, 1298–1309.
- [17] Laakso M, Kuusisto J. Insulin resistance and hyperglycaemia in cardiovascular disease development. *Nat Rev Endocrinol*. 2014 May;10(5):293-302. doi: 10.1038/nrendo.2014.29.
- [18] Laakso, M. et al. Asymptomatic atherosclerosis and insulin resistance. *Arterioscler. Thromb.* 11, 1068–1076 (1991).
- [19] Bressler, P., Bailey, S. R., Matsuda, M. DeFronzo, R. A. Insulin resistance and coronary artery disease. *Diabetologia* 39, 1345–1350 (1996).
- [20] Laakso, M. Cardiovascular disease in type 2 diabetes: challenge for treatment and prevention. *J. Intern. Med.* 249, 225–235 (2001).
- [21] Laakso, M. Insulin resistance and coronary heart disease. *Curr. Opin. Lipidol.* 7, 217–226 (1996).
- [22] Gast, K. B., Tjeerdema, N., Stijnen, T., Smit, J. W. Dekkers, O. M. Insulin resistance and risk of incident cardiovascular events in adults without diabetes: meta-analysis. *PLoS ONE* 7, e52036 (2012).

- [23] Eddy, D., Schlessinger, L., Kahn, R., Peskin, B. Schiebinger, R. Relationship of insulin resistance and related metabolic variables to coronary artery disease: a mathematical analysis. *Diabetes Care* 32, 361–366 (2009).
- [24] Steinberg, H. O., Brechtel, G., Johnson, A., Fineberg, N. Baron, A. D. Insulin-mediated skeletal muscle vasodilation is nitric oxide dependent. A novel action of insulin to increase nitric oxide release. *J. Clin. Invest.* 94, 1172–1179 (1994).
- [25] Steinberg, H. O. et al. Obesity/insulin resistance is associated with endothelial dysfunction. Implications for the syndrome of insulin resistance. *J. Clin. Invest.* 97, 2601–2610 (1996).
- [26] Bornfeldt, K. E. Tabas, I. Insulin resistance, hyperglycemia, and atherosclerosis. *Cell Metab.* 14, 575–585 (2011).
- [27] Myoishi, M. et al. Increased endoplasmic reticulum stress in atherosclerotic plaques associated with acute coronary syndrome. *Circulation* 116, 1226–1233 (2007).
- [28] Moore, K. J. Tabas, I. Macrophages in the pathogenesis of atherosclerosis. *Cell* 145, 341–355 (2011).
- [29] Sharif S, Groenwold RHH, van der Graaf Y, Berkelmans GFN, Cramer MJ, Visseren FLJ, Westerink J; SMART study group. Mediation analysis of the relationship between type 2 diabetes and cardiovascular events and all-cause mortality: Findings from the SMART cohort. *Diabetes Obes Metab.* 2019 Aug;21(8):1935-1943. doi: 10.1111/dom.13759.
- [30] Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. *Am J Epidemiol.* 2012; 176:190-195.
- [31] Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. *Am J Epidemiol.* 2014;179: 513-518.
- [32] Hossin MZ, Koupil I, Falkstedt DEarly life socioeconomic position and mortality from cardiovascular diseases: an application of causal mediation analy-

- sis in the Stockholm Public Health Cohort *BMJ Open* 2021;9:e026258. doi: 10.1136/bmjopen-2018-026258
- [33] Burt VL, Harris T. The third National Health and Nutrition Examination Survey: contributing data on aging and health. *Gerontologist*. 1994 Aug;34(4):486-90. doi: 10.1093/geront/34.4.486.
- [34] Gunter E, Lewis B and Koncikowski S. Laboratory Procedures Used for the Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. 1996.
- [35] Martin SS, Blaha MJ, Elshazly MB, et al.. Comparison of a novel method vs the Friedewald equation for estimating low-density lipoprotein cholesterol levels from the standard lipid profile. *JAMA*. 2013;310(19):2061-2068. doi: 10.1001/jama.2013.280532
- [36] Sampson M, Ling C, Sun Q, Harb R, Ashmaig M, Warnick R, Sethi A, Fleming JK, Otvos JD, Meeusen JW, Delaney SR, Jaffe AS, Shamburek R, Amar M, Remaley AT. A New Equation for Calculation of Low-Density Lipoprotein Cholesterol in Patients With Normolipidemia and/or Hypertriglyceridemia. *JAMA Cardiol*. 2020 May 1;5(5):540-548. doi: 10.1001/jamacardio.2020.0013.
- [37] Helmut Spaeth (1991) *Mathematical Algorithms for Linear Regression*, Academic Press, pp. 304.
- [38] Marill KA. Advanced statistics: linear regression, part II: multiple linear regression. *Acad Emerg Med*. 2004 Jan;11(1):94-102. doi: 10.1197/j.aem.2003.09.006.
- [39] Hand, D.J., Daly, F. , Lunn, A.D., McConway, K., and Ostrowski E. (1994) *The Handbook of Small Data Sets*. Chapman and Hall, London.
- [40] Pintilie M. Análisis de riesgos competitivos [An introduction to competing risks analysis]. *Rev Esp Cardiol*. 2011 Jul;64(7):599-605. Spanish. doi: 10.1016/j.recesp.2011.03.017. Epub 2011 May 31. PMID: 21621892.

- [41] Alba-Gutiérrez, S. Modelos de riesgos competitivos; proyectos fin de master. Master oficial en Estadística Aplicada, Universidad de Granada. Disponible en: [https://masteres.ugr.es/moea/pages/tfm-1112/memoria\\_susana\\_alba/!](https://masteres.ugr.es/moea/pages/tfm-1112/memoria_susana_alba/)
- [42] David, H.A. and Moeschberger, M.L. (1978) *The Theory of Competing Risks*. Griffin, London.
- [43] Porta, N., Gómez, G., Calle, M.L. y Malats, N. (2007). *Competing risks methods*. Technical Report DR 2007/14
- [44] Rebas P. Conceptos básicos del análisis de supervivencia [Basic concepts in survival analysis]. *Cir Esp.* 2005 Oct;78(4):222-30. Spanish. doi: 10.1016/s0009-739x(05)70923-4
- [45] Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, 2(1), 85-97.
- [46] Crowder, M.J. (2001). *Classical competing risks*. Chapman Hall/CRC: Boca raton.
- [47] Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N., Farewell, V.T. y Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks, *Biometrics*, 34: 541-554
- [48] Abd ElHafeez, Samar et al. "Methods to Analyze Time-to-Event Data: The Cox Regression Analysis." *Oxidative medicine and cellular longevity* vol. 2021 1302811. 30 Nov. 2021, doi:10.1155/2021/1302811
- [49] George, Andrew et al. "What's the Risk: Differentiating Risk Ratios, Odds Ratios, and Hazard Ratios?." *Cureus* vol. 12,8 e10047. 26 Aug. 2020, doi:10.7759/cureus.10047
- [50] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Amer Statist Assoc.* 1999;94:496–509

- [51] Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170(2):244-256. <https://doi.org/10.1093/aje/kwp107>
- [52] Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389-2430. <https://doi.org/10.1002/sim.2712>
- [53] Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med*. 2017;36(27):4391-4400. doi:10.1002/sim.7501
- [54] Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol*. 2013;66(6):648-653.
- [55] Sharif S, Groenwold RHH, van der Graaf Y, Berkelmans GFN, Cramer MJ, Visseren FLJ, Westerink J; SMART study group. Mediation analysis of the relationship between type 2 diabetes and cardiovascular events and all-cause mortality: Findings from the SMART cohort. *Diabetes Obes Metab*. 2019 Aug;21(8):1935-1943. doi: 10.1111/dom.13759.
- [56] Zhang Z, Zheng C, Kim C, Van Poucke S, Lin S, Lan P. Causal mediation analysis in the context of clinical research. *Ann Transl Med*. 2016 Nov;4(21):425. doi: 10.21037/atm.2016.11.11.
- [57] MacKinnon DP, Pirlott AG. Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Pers Soc Psychol Rev* 2015;19:30-43.
- [58] Linden A, Karlson KB. Using mediation analysis to identify causal mechanisms in disease management interventions. *Health Serv Outcomes Res Method* 2013;13:86-108.
- [59] VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology*. 2011;22(4):582-585. doi:10.1097/EDE.0b013e31821db37e

- [60] Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011 Jul;22(4):575-81. doi: 10.1097/EDE.0b013e31821c680c. PMID: 21552129.
- [61] Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. In: *Proceedings of the International Joint Conferences on Artificial Intelligence*; 2005:357:363.
- [62] VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2:457-468.
- [63] Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Eval Rev*. 1981;5(5):602-619
- [64] Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986;51(6):1173-1182.
- [65] VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis with a dichotomous outcome. *American Journal of Epidemiology*. 2010;172:1339-1348.
- [66] Tein J-Y, MacKinnon DP. Estimating mediated effects with survival data. In: Yanai H, Rikkyo AO, Shigemasu K, Kano Y, Meulman JJ, editors. *New Developments on Psychometrics*. Springer-Verlag Tokyo Inc; Tokyo, Japan: 2003. pp. 405-412.
- [67] Cho SH, Huang YT. Mediation analysis with causally ordered mediators using Cox proportional hazards model. *Stat Med*. 2019 Apr 30;38(9):1566-1581. doi: 10.1002/sim.8058.
- [68] Huang YT, Yang HI. Causal Mediation Analysis of Survival Outcome with Multiple Mediators. *Epidemiology*. 2017 May;28(3):370-378. doi: 10.1097/EDE.0000000000000651.
- [69] Freund Y, Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. 1997, 55, 119-139.

- [70] Chollet F, Allaire JJ. Deep learning with R. 2018; Shelter Island, NY: Manning Publications.
- [71] Whittle R, Mansell G, Jellema P, van der Windt D. Applying causal mediation methods to clinical trial data: What can we learn about why our interventions (don't) work? *Eur J Pain*. 2017 Apr;21(4):614-622. doi: 10.1002/ejp.964. Epub 2016 Oct 14.
- [72] Di Serio C, Rinott Y, Scarsini M. Simpson's paradox in survival models. *Scand J Stat*. 2009;36(3):463-80.
- [73] Sjolander A, Dahlqvist E, Zetterqvist J. A note on the noncollapsibility of rate differences and rate ratios. *Epidemiology*. 2016;27(3):356-9.
- [74] Kato T, Inoue T, Inagaki H, Hashimoto S, Hikichi Y, Tanaka A, Isobe M, Node K. Remnant-like lipoprotein particle level and insulin resistance are associated with in-stent restenosis in patients with stable angina. *Coron Artery Dis*. 2007 Jun;18(4):319-22. doi: 10.1097/MCA.0b013e32802c7cc4
- [75] Nakajima K, Tokita Y, Tanaka A. Hypothesis: Postprandial remnant lipoproteins are the causal factors that induce the insulin resistance associated with obesity. *Clin Chim Acta*. 2018 Oct;485:126-132. doi: 10.1016/j.cca.2018.06.029.
- [76] J.R.I. Goudriaan, P.J. Tacke, V.E. Dahlmans, M.J. Gijbels, K.W. van Dijk, L.M. Havekes, et al., Protection from obesity in mice lacking the VLDL receptor, *Arterioscler. Thromb. Vasc. Biol*. 21 (2001) 1488-1493.
- [77] S. Takahashi, Triglyceride rich lipoprotein -LPL-VLDL receptor and Lp(a)-VLDL receptor pathways for macrophage foam cell formation, *J. Atheroscler. Thromb*. 24 (2017) 552-559.
- [78] Liang, Y., Vetrano, D.L. Qiu, C. Serum total cholesterol and risk of cardiovascular and non-cardiovascular mortality in old age: a population-based study. *BMC Geriatr* 17, 294 (2017). <https://doi.org/10.1186/s12877-017-0685-z>

- [79] Newson RS, Felix JF, Heeringa J, Hofman A, Witteman JC, Tiemeier H. Association between serum cholesterol and noncardiovascular mortality in older age. *J Am Geriatr Soc.* 2011 Oct;59(10):1779-85. doi: 10.1111/j.1532-5415.2011.03593.x.
- [80] Mortensen MB, Nordestgaard BG. Elevated LDL cholesterol and increased risk of myocardial infarction and atherosclerotic cardiovascular disease in individuals aged 70-100 years: a contemporary primary prevention cohort. *Lancet.* 2020 Nov 21;396(10263):1644-1652. doi: 10.1016/S0140-6736(20)32233-9.
- [81] Vladu IM, Forțofoiu M, Clenciu D, Forțofoiu MC, Pădureanu R, Radu L, Cojan ȘTT, Rădulescu PM, Pădureanu V. Insulin resistance quantified by the value of HOMA-IR and cardiovascular risk in patients with type 2 diabetes. *Exp Ther Med.* 2022 Jan;23(1):73. doi: 10.3892/etm.2021.10996.
- [82] Kang, E.S.; Yun, Y.S.; Park, S.W.; Kim, H.J.; Ahn, C.W.; Song, Y.D.; Cha, B.S.; Lim, S.K.; Kim, K.R.; Lee, H.C. Limitation of the validity of the homeostasis model assessment as an index of insulin resistance in Korea. *Metabolism* 2005, 54, 206–211
- [83] Luo P, Cao Y, Li P, Li W, Song Z, Fu Z, Zhou H, Yi X, Zhu L, Zhu S. TyG Index Performs Better Than HOMA-IR in Chinese Type 2 Diabetes Mellitus with a BMI < 35 kg/m<sup>2</sup>: A Hyperglycemic Clamp Validated Study. *Medicina (Kaunas).* 2022 Jun 30;58(7):876. doi: 10.3390/medicina58070876.
- [84] Lapointe-Shaw, L., Bouck, Z., Howell, N.A. et al. Mediation analysis with a time-to-event outcome: a review of use and reporting in healthcare research. *BMC Med Res Methodol* 18, 118 (2018). <https://doi.org/10.1186/s12874-018-0578-7>
- [85] Quispe R, Martin SS, Michos ED, Lamba I, Blumenthal RS, Saeed A, Lima J, Puri R, Nomura S, Tsai M, Wilkins J, Ballantyne CM, Nicholls S, Jones SR, Elshazly MB. Remnant cholesterol predicts cardiovascular disease beyond LDL and ApoB: a primary prevention study. *Eur Heart J.* 2021 Nov 7;42(42):4324-4332. doi: 10.1093/eurheartj/ehab432

- [86] Rosenson RS, Davidson MH, Hirsh BJ, Kathiresan S, Gaudet D. Genetics and causality of triglyceride-rich lipoproteins in atherosclerotic cardiovascular disease. *J Am Coll Cardiol.* 2014;64(23):2525–2540.
- [87] Sandesara PB, Virani SS, Fazio S, Shapiro MD. The Forgotten Lipids: Triglycerides, Remnant Cholesterol, and Atherosclerotic Cardiovascular Disease Risk. *Endocr Rev.* 2019 Apr 1;40(2):537-557. doi: 10.1210/er.2018-00184.

# Anexo: Código de R

A continuación, se anexa en extenso el código utilizado. El código y su renderización se encuentran también disponibles en un repositorio digital: [https://github.com/Vargas-Vazquez/Tesina\\_EEA\\_2022/blob/22e59af19e4901e6c47fbf70cc2722550da96ad9/C%C3%B3digo\\_an%C3%A1lisis](https://github.com/Vargas-Vazquez/Tesina_EEA_2022/blob/22e59af19e4901e6c47fbf70cc2722550da96ad9/C%C3%B3digo_an%C3%A1lisis).

```
####---- Database management ----####
pacman::p_load(readxl, haven, tidyverse, ggpubr, lmtest, nortest, gtools,
               data.table, caret, glmnet, survival, flextable, blandr,
               BlandAltmanLeh,
               rms, bestNormalize, flexsurv, pROC, timeROC, fmsb, factoextra,
               gridExtra, nhanesA, wesanderson, forestmodel, ggedit,
               FactoMineR, fpc, NbClust, ggimage, glmnet, ggsci, survminer,
               cluster, ggplotify, UpSetR,
               nortest, viridis, officer, magrittr, mice,
               PerformanceAnalytics, corrplot, mice, performance)

### Loading databases #####
#NHANES-III
NHANES3<-fread("nhanes3.csv", na = c("", "N/A", "NA", "na", "#N/A", "8888",
                                   "8888", "888888", "888", "5555", "999", "9
                                   "99999", "999999", "9998"))
mortalidad3<-fread("nhanes3_mort.csv"); mortalidad3$SEQN<-mortalidad3$seqn
```

```

### NHANES-III management ###
NHANES0 <- NHANES3 %>% filter(HSAGEIR.x>=20)%>%
  mutate("BMI"=BMPBMI, "Thigh_circumference"=BMPTHICI,
         "Arm_circumference"=BMPARMC,
         "Weight"= BMPWT,"Height"=BMPHT, "Waist"=BMPWAIST,
         "Triceps_skinfold"=BMPTRI,
         "Subscapular_skinfold"=BMPSUB, "Leg_length"=BMPLEG,
         "Arm_length"=BMPARML, "ICE"=BMPWAIST/BMPHT, "Diabetes"=HAD1,
         "Hypertension"=HAE2, "Asthma"=HAC1E, "Arthritis"=HAC1A,
         "Heart_failure"=HAC1C, "Heart_attack"=HAF10, "Stroke"=HAC1D,
         "Emphysema"=HAC1G, "Bronchitis"=HAC1F, "Malignancy"=HAC1O,
         "Sex"=factor(HSSEX.x, levels= c(1,2), labels= c("Men", "Women")),
         "Age"=HSAGEIR.x, "Ethnicity"=DMARETHN.x, "HDL"=HDP, "TG"=TGP,
         "ApoB"=ABP, "Cholesterol"=TCP, "Glucose"=SGP, "HbA1c"=GHP,
         "C_peptide"=C1PSI, "Insulin"=I1P,
         "METS_IR"=(log((2*SGP)+TGP)*BMPBMI)/(log(HDP)), no_HDL=TCP-HDP,
         "LDL_S"=(TCP/0.948)-(HDP/0.971)-((TGP/8.56)+((TGP*no_HDL)/2140)-
                                         ((TGP*TGP)/16100))-9.4,
         "Remanentes"=TCP-HDP-LDL_S, "VLDL_C"=((TGP/8.59)+((TGP*no_HDL)/2250
                                         ((TGP*TGP)/16500))),
         "HOMA2IR"='HOMA2 IR', "HOMA2B"='HOMA2 %B', "HOMA2S"='HOMA2 %S',
         "Fuma_cig"=HAR3,"Fuma_pur"=HAR24, "Fuma_pip"=HAR27, "creatinina"=CE

write_csv(NHANES0, "nhanes3_basal.csv")
#Race/Ethnicity: 1=White, 2=Black, 3=Mexican-American, 4=Other
NHANES0$Remanentes
NHANES0$LDL_S
NHANES0$VLDL_C
NHANES0$creatinina

#Cause-specific mortality

```

```

NHANES0 <- merge(NHANES0,mortalidad3,by="SEQN")
d1<-dummies::dummy(NHANES0$ucod_leading)
colnames(d1)<-c("Heart_diseases","Malignant_neoplasms",
               "Chronic_lower_respiratory_diseases","Accidents",
               "Cerebrovascular_diseases","Alzheimer_disease",
               "Diabetes_mellitus","Influenza_or_pneumonia",
               "Nephone_diseases","Other_causes","Alive"); NHANES0<-cbind(N
nhanes<-NHANES0 %>% dplyr::select(SEQN, Age, Sex, Ethnicity, mortstat,permth
                               permth_exm,BMI, Thigh_circumference,
                               Arm_circumference, Triceps_skinfold,
                               Subscapular_skinfold,Leg_length,
                               Arm_length, Waist, Height, Weight, ICE,HDL
                               Cholesterol, ApoB, Glucose, HbA1c, C_pepti
                               Insulin, METS_IR, no_HDL, LDL_S, Remanente
                               VLDL_C,HOMA2IR, HOMA2B, HOMA2S,"Heart_dise
                               "Malignant_neoplasms",
                               "Chronic_lower_respiratory_diseases","Acci
                               "Cerebrovascular_diseases","Alzheimer_dise
                               "Diabetes_mellitus","Influenza_or_pneumoni
                               "Nephone_diseases","Other_causes","Alive",
                               Diabetes, Hypertension, Asthma, Arthritis,
                               Heart_failure,
                               Heart_attack, Stroke, Emphysema, Bronchiti
                               Malignancy, Fuma_cig, Fuma_pip, Fuma_pur,
                               creatinina); nhanes$id <- rep(1,nrow(nhane

#Comorbidities
nhanes$Diabetes<-ifelse(nhanes$Diabetes==1, 1, 0);
nhanes$Asthma<-ifelse(nhanes$Asthma==1, 1, 0)
nhanes$Arthritis<-ifelse(nhanes$Arthritis==1, 1, 0);
nhanes$Heart_failure<-ifelse(nhanes$Heart_failure==1, 1, 0)

```

```
nhanes$Heart_attack<-ifelse(nhanes$Heart_attack==1, 1, 0);
nhanes$Emphysema<-ifelse(nhanes$Emphysema==1, 1, 0)
nhanes$Bronchitis<-ifelse(nhanes$Bronchitis==1, 1, 0);
nhanes$Malignancy<-ifelse(nhanes$Malignancy==1, 1, 0)
nhanes$Stroke<-ifelse(nhanes$Stroke==1, 1, 0);
nhanes$Hypertension<-ifelse(nhanes$Hypertension==1, 1, 0)

#Number of comorbidities
nhanes$num_comorb<-nhanes$Diabetes+nhanes$Asthma+nhanes$Arthritis+
  nhanes$Heart_failure+nhanes$Heart_attack+nhanes$Emphysema+
  nhanes$Bronchitis+nhanes$Malignancy+nhanes$Stroke+nhanes$Hypertension

#Race/Ethnicity
nhanes$Ethnicity<-factor(nhanes$Ethnicity, labels = c("White", "Black",
                                                    "Mexican-American", "0

#tabaquismo
nhanes$Fuma_cig<-ifelse(nhanes$Fuma_cig==1, 1, 0);
nhanes$Fuma_pip<-ifelse(nhanes$Fuma_pip==1, 1, 0);
nhanes$Fuma_pur<-ifelse(nhanes$Fuma_pur==1, 1, 0)

write_csv(nhanes, "nhanes3_mortalidad.csv")

##### Descriptivos #####

## Selecci n de la base ####

base<- read_excel("nhanes3_mortalidad_F.xlsx")

### Valores perdidos ###
```

```

sapply(base , function(x) sum(is.na(x)))
base1 <- base[!is.na(base$LDL_S),]
base1 <- base1[!is.na(base1$METS_IR),]
base1 <- base1[!is.na(base1$mortstat),]
sapply(base1, function(x) sum(is.na(x)))

#Eliminar variables que no se requieren para el analisis #

base<-select(base1, -Thigh_circumference, -Arm_circumference, -Triceps_skinf
             -Subscapular_skinfold, -Leg_length, -Arm_length )
sapply(base , function(x) sum(is.na(x)))

##### Imputaci n de datos #####
### Removemos Apob de la imputaci n
base2<-select(base , -ApoB)
set.seed(123456)
imp<-mice(base2, m=1, maxit=1)
x1<-complete(imp, "long")[,-c(1:3)]
base<-as.data.frame(x1)

### Creaci n de variable mortalidad cardiovascular ateroscler tica; ###
#Heart diseases and Cerebrovascular diseases #

base$ASCVD<-NULL
base$ASCVD[base$Heart_diseases==1 & base$Cerebrovascular_diseases==1]<-1
base$ASCVD[base$Heart_diseases==0 & base$Cerebrovascular_diseases==1]<-1
base$ASCVD[base$Heart_diseases==1 & base$Cerebrovascular_diseases==0]<-1
base$ASCVD[base$Heart_diseases==0 & base$Cerebrovascular_diseases==0]<-0

base$ASCVD_mort <- NULL

```

```
base$ASCVD_mort [base$ASCVD==1 & base$mortstat==1]<-1
base$ASCVD_mort [base$ASCVD==0 & base$mortstat==1]<-2
base$ASCVD_mort [base$ASCVD==0 & base$mortstat==0]<-0
base$ASCVD_mort<-factor(base$ASCVD_mort, labels = c("Censored",
                                                    "ASCVD_mortality", "Other"))

##### Segmentamos la base; base_dm y base_Nodm #####

#Creamos las variables para DM
base$dm_hba1c [base$HbA1c>=6.5]<-1; base$dm_hba1c [base$HbA1c<6.5]<-0
base$dm_G [base$Glucose>=126]<-1; base$dm_G [base$Glucose<126]<-0

# Variable dm general de diabetes; DMF#

base$DM<-NULL
base$DM [base$Diabetes==1 & base$dm_hba1c==1]<-1
base$DM [base$Diabetes==0 & base$dm_hba1c==1]<-1
base$DM [base$Diabetes==1 & base$dm_hba1c==0]<-1
base$DM [base$Diabetes==0 & base$dm_hba1c==0]<-0

base$DMF<-NULL
base$DMF [base$DM==1 & base$dm_G==1]<-1
base$DMF [base$DM==0 & base$dm_G==1]<-1
base$DMF [base$DM==1 & base$dm_G==0]<-1
base$DMF [base$DM==0 & base$dm_G==0]<-0

### tabaquismo
base$cigarros<-NULL
base$cigarros [base$Fuma_cig==1 & base$Fuma_pur==1]<-1
base$cigarros [base$Fuma_cig==0 & base$Fuma_pur==1]<-1
base$cigarros [base$Fuma_cig==1 & base$Fuma_pur==0]<-1
base$cigarros [base$Fuma_cig==0 & base$Fuma_pur==0]<-0
```

```

base$Tab<-NULL
base$Tab[base$cigarros==1 & base$Fuma_pip==1]<-1
base$Tab[base$cigarros==0 & base$Fuma_pip==1]<-1
base$Tab[base$cigarros==1 & base$Fuma_pip==0]<-1
base$Tab[base$cigarros==0 & base$Fuma_pip==0]<-0

write_sav(base, "nhanes_mortalidad.sav")
#Segmetamos bases #
base_nodm<-base%>%filter(DMF==0)
base_dm<-base%>%filter(DMF==1)

### C lculo de incidencia por casa 1000 personas a os #####
## mortalidad en el total de la poblaci n
mort<- base%>%group_by(mortstat) %>%
  summarise(cases=n(), time=sum(base$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$mortstat==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

## mortalidad cardiovascular en el total de la poblaci n
mort<- base%>%group_by(ASCVD_mort) %>%
  summarise(cases=n(), time=sum(base$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$ASCVD_mort=="ASCVD_mortality"]
lambda<-(cases/time)*1000

```

```
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# mortalidad general
#para px sin dm
mort<- base_nodm%>%group_by(mortstat) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$mortstat==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(mortstat) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$mortstat==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# mortalidad general cardiovascular
# sin dm
mort<- base_nodm%>%group_by(ASCVD_mort) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
```

```

time<- sum(mort$time)
cases<- mort$cases[mort$ASCVD_mort=="ASCVD_mortality"]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(ASCVD_mort) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$ASCVD_mort=="ASCVD_mortality"]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

## mortalidad por diabetes
# sin dm
mort<- base_nodm%>%group_by(Diabetes_mellitus) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Diabetes_mellitus==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Diabetes_mellitus) %>%

```

```
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Diabetes_mellitus==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

## mortalidad por neoplasias
# sin dm
mort<- base_nodm%>%group_by(Malignt_neoplasms) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Malignt_neoplasms==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Malignt_neoplasms) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Malignt_neoplasms==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)
```

```

# mortalidad por EPOC
# sin dm
mort<- base_nodm%>%group_by(Chronic_lower_respiratory_diseases) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Chronic_lower_respiratory_diseases==1]
lambda<-(cases/time)*1000
lambda.se<-sqrt(cases/time^2)*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Chronic_lower_respiratory_diseases) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Chronic_lower_respiratory_diseases==1]
lambda<-(cases/time)*1000
lambda.se<-sqrt(cases/time^2)*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

## mortalidad por accidentes
# sin dm
mort<- base_nodm%>%group_by(Accidents) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Accidents==1]
lambda<-(cases/time)*1000
lambda.se<-sqrt(cases/time^2)*1000; LI<-lambda-(1.96*lambda.se);

```

```
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Accidents) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Accidents==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

### mortalidad por alzheimer
# sin dm
mort<- base_nodm%>%group_by(Alzheimer_disease) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Alzheimer_disease==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Alzheimer_disease) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Alzheimer_disease==1]
```

```

lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

## mortalidad por pnemonia
# sin dm
mort<- base_nodm%>%group_by(Influenza_or_pneumonia) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Influenza_or_pneumonia==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Influenza_or_pneumonia) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Influenza_or_pneumonia==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

### mortalidad por enfermedad renal
# sin dm
mort<- base_nodm%>%group_by(Nephone_diseases) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))

```

```
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Nephone_diseases==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
mort<- base_dm%>%group_by(Nephone_diseases) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Nephone_diseases==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

##3 mortalidad por otra causa
# sin dm
mort<- base_nodm%>%group_by(Other_causes) %>%
  summarise(cases=n(), time=sum(base_nodm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Other_causes==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

# con dm
```

```

mort<- base_dm%>%group_by(Other_causes) %>%
  summarise(cases=n(), time=sum(base_dm$permth_int/12))
#Luego extraes el tiempo de cada caso
time<- sum(mort$time)
cases<- mort$cases[mort$Other_causes==1]
lambda<-(cases/time)*1000
lambda.se<-(sqrt(cases/time^2))*1000; LI<-lambda-(1.96*lambda.se);
LS<-lambda+(qnorm(1-0.05/2)*lambda.se)
c(lambda, LI, LS)

#### Distribuci n de las variables de inter s ####
#sujetos sin diabetes
bas_1<- base_nodm%>%dplyr::select(Age, BMI, ICE, TG,Cholesterol, LDL_S,
                                Remanentes,Rem_M,
                                no_HDL, HDL, HOMA2IR)

chart.Correlation(bas_1, histogram=TRUE, pch=19, method="spearman")
## se eliminaron del analisis a los sujetos con TG mayores a 800 y HOMA2IR<
base_nodm<-base_nodm%>%filter(TG<800)
base_nodm<-base_nodm%>%filter(HOMA2IR<15)
## se transforman las variables que no muestran una distribuci n normal
base_nodm$Rem_Sn<-predict(bestNormalize(base_nodm$Remanentes));
#bestNormalize(base_nodm$Remanentes)
base_nodm$Rem_Mn<-predict(bestNormalize(base_nodm$Rem_M));
#bestNormalize(base_nodm$Rem_M)
base_nodm$HOMA2IR_n<-predict(bestNormalize(base_nodm$HOMA2IR));
#bestNormalize(base_nodm$HOMA2IR)
base_nodm$BMI_n<-predict(bestNormalize(base_nodm$BMI));
#bestNormalize(base_nodm$BMI)
base_nodm$age_n<-predict(bestNormalize(base_nodm$Age));
#bestNormalize(base_nodm$Age)

```

```
## se evalu la distribuci n despues de normalizar
bas_1.2<- base_nodm%>%dplyr::select(age_n, BMI_n, TG,Cholesterol,
                                   Rem_Sn,Rem_Mn,
                                   no_HDL, HDL, HOMA2IR_n)
chart.Correlation(bas_1.2, histogram=TRUE, pch=19, method="spearman")

#sujetos con diabetes
bas_2<- base_dm%>%dplyr::select(Age, BMI, ICE, TG,Cholesterol,
                                   Rem_S,Rem_M,
                                   no_HDL, HDL, HOMA2IR)

chart.Correlation(bas_2, histogram=TRUE, pch=19, method="spearman")

## se eliminaron del analisis a los sujetos con TG mayores a 800 y HOMA2IR<
base_dm<-base_dm%>%filter(TG<800)
quantile(base_dm$HOMA2IR, p=0.99);summary(base_dm$HOMA2IR)
base_dm<-base_dm%>%filter(HOMA2IR<20)
## se transforman las variables que no muestran una distribuci n normal
base_dm$Rem_Sn<-predict(bestNormalize(base_dm$Rem_S))
#bestNormalize(base_dm$Rem_S)
base_dm$Rem_Mn<-predict(bestNormalize(base_dm$Rem_M))
#bestNormalize(base_dm$Rem_M)
base_dm$HOMA2IR_n<-predict(bestNormalize(base_dm$HOMA2IR))
#bestNormalize(base_dm$HOMA2IR)
base_dm$age_n<-predict(bestNormalize(base_dm$Age))
#bestNormalize(base_dm$Age)

## se evalu la distribuci n despues de normalizar
bas_2.1<- base_dm%>%dplyr::select(age_n, BMI, ICE, TG,Cholesterol,
                                   Rem_Sn,Rem_Mn,
                                   no_HDL, HDL, HOMA2IR_n)
chart.Correlation(bas_2.1, histogram=TRUE, pch=19, method="spearman")
```

```
##### Modelo de RLM en pacientes sin diabetes mellitus#####

#Remanentes calculados con Sampson#
fig_1<-ggplot(base_nodm, aes(x=HOMA2IR_n, y=Rem_Sn))+geom_point()+
  geom_smooth(method = "lm")+labs(x="HOMA2IR", y="OrderNorm(Remanentes)")+
  theme_classic()

base_nodm$Edad<-base_nodm$age_n
base_nodm$HOMA2_IR<-base_nodm$HOMA2IR_n
base_nodm$Etn<-base_nodm$Ethnicity
base_nodm$Comorb<-base_nodm$num_comorb
base_nodm$IMC<-base_nodm$BMI_n

# modelo para remanentes calculados con Sampson #
m_S<-lm(Rem_Sn~HOMA2IR_n+age_n+Sex+Ethnicity+BMI_n+num_comorb+LDLC_S+Tab, ba
summary(m_S)
BIC(m_S)
#m_S<-lm(Rem_Sn~HOMA2_IR+Edad+Sex+Etn+IMC+Comorb+LDLC_S+Tab, base_nodm)
summary(m_S)
check_model(m_S)

# homoscedasticidad de la varianza
bptest(m_S, data=base_nodm, studentize = T)
library(skedastic)
skedastic::white_lm(m_S)
## autocorrelacion de los errores
bgtest(m_S)
# normalidad de los errores
ad.test(m_S$residuals)
#m todo de Hotelling-Shef
library(multcomp)
```

```
K <- cbind(diag(length(coef(m_S))))
rownames(K) <- names(coef(m_S))
sim <- glht(m_S, linfct = K)
confint(sim, level = 0.95)

#Remanentes calculados con Martin#

fig_2<-ggplot(base_nodm, aes(x=HOMA2IR_n, y=Rem_Mn))+geom_point()+
  geom_smooth(method = "lm")+
  labs(x="HOMA2IR", y="OrderNorm(Remanentes_□(M))")+
  theme_classic()

# Modelo
m_M<-lm(Rem_Mn~HOMA2IR_n+age_n+Sex+Ethnicity+BMI_n+num_comorb+LDL_M+Tab, bas
summary(m_M)
BIC(m_M);BIC(m_S)
#m_M<-lm(Rem_Mn~HOMA2_IR+Edad+Sex+Etn+IMC+Comorb+LDLC_S+Tab, base_nodm)
check_model(m_M)

# homoscedasticidad de la varianza
bptest(m_M, data=base_nodmm, studentize = T)
## autocorrelacion de los errores
bgttest(m_M)
# normalidad de los errores
ad.test(m_M$residuals)
ks.test(m_M$residuals, "pnorm")
#m todo de Hotelling-Shef
K <- cbind(diag(length(coef(m_M))))
rownames(K) <- names(coef(m_M))
sim <- glht(m_M, linfct = K)
```

```

confint(sim, level = 0.95)

### Modelos de RLM en pacientes con DM ###
base_dm$black[base_dm$Ethnicity=="Black"]<-1;
base_dm$black[base_dm$Ethnicity!="Black"]<-0
base_dm$sexo[base_dm$Sex=="Women"]<-0
base_dm$sexo[base_dm$Sex!="Women"]<-1
library(nephro)
# calculo de TFG

base_dm$TFG<-CKDEpi.creat(base_dm$CR_S, base_dm$sexo, base_dm$Age, base_dm$b
summary(base_dm$TFG)
#Sampson
base_dm<-base_dm%>%filter(Remanentes>=0)
Fig_3<-ggplot(base_dm, aes(x=HOMA2IR_n, y=Rem_Sn))+geom_point()+
  geom_smooth(method="lm")+labs(x="HOMA2IR", y="OrderNorm(Remanentes_(S))")+
  theme_classic()

m_Sdm<-lm(Rem_Sn~HOMA2IR_n+age_n+Sex+Ethnicity+BMI+num_comorb+LDLC_S+Tab, ba
summary(m_Sdm)
##
base_dm$Edad<-base_dm$age_n
base_dm$HOMA2_IR<-base_dm$HOMA2IR_n
base_dm$Etn<-base_dm$Ethnicity
base_dm$Comorb<-base_dm$num_comorb
base_dm$IMC<-base_dm$BMI

#m_Sdm<-lm(Rem_Sn~HOMA2_IR+Edad+Sex+Etn+IMC+Comorb+LDLC_S+Tab, base_dm)
check_model(m_Sdm)

#m_S<-lm(Rem_Sn~HbA1c+Age+sexo+etni+BMI_n+num_comorb, base_dm)
#summary(m_S)

```

```

#check_model(m_S)

# homoscedasticidad de la varianza

bptest(m_Sdm, data=base_dm)

#Autocorrelación de los errores
bgttest(m_Sdm)
# normalidad de los errores
ad.test(m_Sdm$residuals)
#m todo de Hotelling-Shef
K <- cbind(diag(length(coef(m_Sdm))))
rownames(K) <- names(coef(m_Sdm))
sim <- glht(m_Sdm, linfct = K)
confint(sim, level = 0.95)

# remanentes calculados con Martin#

Fig_4<-ggplot(base_dm, aes(x=HOMA2IR_n, y=Rem_Mn))+geom_point()+
  geom_smooth(method="lm")+labs(x="HOMA2IR", y="OrderNorm(Remanentes(M))")+
  theme_classic()

#modelo
m_Mdm<-lm(Rem_Mn~HOMA2IR_n+Age+Sex+Ethnicity+BMI+num_comorb+LDL_M+Tab, base_
summary(m_Mdm)
#m_Mdm<-lm(Rem_Mn~HOMA2_IR+Edad+Sex+Etn+IMC+Comorb+LDLC_S+Tab, base_dm)
check_model(m_Mdm)

#m_S<-lm(Rem_Mn~HbA1c+Age+sexo+etni+BMI_n+num_comorb, base_dm)
#summary(m_S)
#check_model(m_S)

```

```
# homoscedasticidad de la varianza

bptest(m_Mdm, data=base_dm)

#Autocorrelaci n de los errores
bptest(m_Mdm)
# normalidad de los errores
ad.test(m_Mdm$residuals)

#m todo de Hotelling-Shef
K <- cbind(diag(length(coef(m_Mdm))))
rownames(K) <- names(coef(m_Mdm))
sim <- glht(m_Mdm, linfct = K)
confint(sim, level = 0.95)

BIC(m_Sdm);BIC(m_Mdm)

#### Modelo de regresi n de cox en sujetos sin dm####

#sujetos sin DM#

#Percentila 85 de HOMA2IR#
quantile(base_nodm$HOMA2IR, p=0.85)
base_nodm$HOMAIR_cat[base_nodm$HOMA2IR<2.5]<-0;
base_nodm$HOMAIR_cat[base_nodm$HOMA2IR>=2.5]<-1

#Percentila 50 de remanentes de colesterol
quantile(base_nodm$Rem_S)
base_nodm$RemS_cat[base_nodm$Rem_S<19]<-0;
base_nodm$RemS_cat[base_nodm$Rem_S>=19]<-1

base_nodm$RemS_s<-scale(base_nodm$Rem_S)
```

```

base_nodm$LDL_S<-scale(base_nodm$LDLC_S)
base_nodm$Edad<-base_nodm$Age
base_nodm$HOMA2_IR<-base_nodm$HOMAIR_cat
base_nodm$Etn<-base_nodm$Ethnicity
base_nodm$Comorb<-base_nodm$num_comorb
base_nodm$IMC<-base_nodm$BMI
## modelo de Fin y Gray para mortalidad cardiovascular ateroscler tica
fgdata_cv <- finegray(Surv(permeth_int/12, ASCVD_mort) ~ ., data=base_nodm,
                      na.action=na.pass)

#modelo Cox-fine-gray

m1_s <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemS_s*Age+
              LDL_S+Sex+Comorb+
              +BMI+Tab+Etn,weight=fgwt, data=fgdata_cv)

summary(m1_s)
BIC(m1_s)

#modelo con iteracci n
#m1_s <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat*RemS_cat+Sex+num_
#              #Ethnicity+BMI,weight=fgwt, data=fgdata_cv)
#summary(m1_s)

# probamos supuestos del modelo de regresi n de cox

test.ph<-cox.zph(m1_s);test.ph
ggcoxzph(test.ph)

### test para observaciones influyentes

ggcoxdiagnostics(m1_s , type ="dfbeta" , linear.predictions = F,
                 ggtheme= theme_bw())

```

```

## deviance residuals

ggcoxdiagnostics(m1_s , type ="deviance" , linear.predictions = F,
                 ggtheme= theme_bw())
quantile(base_nodm$RemS_scale)
## simulaci n con la edad
library(simPH)
Sim2 <- coxsimInteract(m1_s, b1 = "RemS_scale", b2 = "Age",
                      X1 = seq(-2, 8, by = 1),
                      X2 = seq(30, 80, by =20),
                      qi = "Hazard_Ratio", ci = 0.95, spin = F)

Fig_simPH_S<-simGG(Sim2, xlab = "Rem-S(scale)", type = "ribbons", lsize = 1,
                  leg.name= "Age", legend="legend", alpha = 0.3, nsim = 100)

## modelos con remanentes calculados con Martin
#Percentila 75 de remanentes de colesterol
quantile(base_nodm$Rem_M, p=0.5)
base_nodm$RemM_cat[base_nodm$Rem_M<21.7]<-0;
base_nodm$RemM_cat[base_nodm$Rem_M>=21.7]<-1
base_nodm$RemM_s<-scale(base_nodm$Rem_M)
base_nodm$LDL_M_s<-scale(base_nodm$LDL_M)

## modelo de Fin y Gray para mortalidad cardiovascular ateroscler tica

fgdata_cv <- finegray(Surv(permeth_int, ASCVD_mort) ~ ., data=base_nodm,
                     na.action=na.pass)
m1_m <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemM_s*Age+Sex+
              Comorb+Etn+BMI+LDL_M_s+Tab,weight=fgwt, data=fgdata_cv)
summary(m1_m)

```

```

BIC(m1_m)
#interaccion#
#m1_m <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat*RemM_cat+
              #Sex+num_comorb+Age+
              #Ethnicity+BMI,weight=fgwt, data=fgdata_cv)
#summary(m1_m)

# probamos supuestos del modelo de regresión de cox

test.phm<-cox.zph(m1_m);test.phm
ggcoxzph(test.phm)
ggcoxzph(test.phm[2])

### test para observaciones influyentes

ggcoxdiagnostics(m1_m , type ="dfbeta" , linear.predictions = F,
                 ggtheme= theme_bw())

## deviance residuals

ggcoxdiagnostics(m1_m , type ="deviance" , linear.predictions = F,
                 ggtheme= theme_bw())

#Simulaciones de la interacción con edad
quantile(base_nodm$RemM_scale)
Sim2 <- coxsimInteract(m1_m, b1 = "RemM_scale", b2 = "Age",
                      X1 = seq(-2, 8, by = 1),
                      X2 = seq(30, 80, by =20),
                      qi = "Hazard_Ratio", ci = 0.95, spin = F)

```

```

Fig_simPH_M<-simGG(Sim2, xlab = "Rem-M1(scale)", type = "ribbons", lsize = 1
                    leg.name= "Age", legend="legend", alpha = 0.8, nsim = 100

### modelos de Fine and Gray en pacientes con diabetes ####
base_dm$cv_mort<-NULL
base_dm$cv_mort[base_dm$Heart_diseases==1 & base_dm$mortstat==1]<-1
base_dm$cv_mort[base_dm$Heart_diseases==0 & base_dm$mortstat==1]<-2
base_dm$cv_mort[base_dm$Heart_diseases==1 & base_dm$mortstat==0]<-1
base_dm$cv_mort[base_dm$Heart_diseases==0 & base_dm$mortstat==0]<-0
base_dm$cv_mort<-factor(base_dm$cv_mort,
                        labels = c("Censored", "cv1mortality", "Other1causes

#Percentila 85 de HOMA2IR#
quantile(base_dm$HOMA2IR, p=0.84)
base_dm$HOMAIR_cat[base_dm$HOMA2IR<4.9]<-0;
base_dm$HOMAIR_cat[base_dm$HOMA2IR>=4.9]<-1

#Percentila 50 de remanentes de colesterol
quantile(base_dm$Rem_S)
base_dm$RemS_cat[base_dm$Rem_S<40]<-0;base_dm$RemS_cat[base_dm$Rem_S>=40]<-1
base_dm$LDLS_scl<-scale(base_dm$LDLC_S)

## modelo de Fin y Gray para mortalidad cardiovascular ateroscler tica
fgdata_cv <- finegray(Surv(permeth_int/12, cv_mort) ~ ., data=base_dm,
                    na.action=na.pass)

## modelo simple
m1_sdm_s<- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemS_cat,
                weight=fgwt, data=fgdata_cv)
summary(m1_sdm_s)

#modelo Cox-fine-gray
m1_sdm <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemS_cat+Age+Sex

```

```
      Comorb+
      Etn+BMI+LDLS_scl+Tab,weight=fgwt, data=fgdata_cv)
summary(m1_sdm)
BIC(m1_sdm)

#modelo con iteracci n
#m1_s <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat*RemS_cat+Sex+
      #num_comorb+Age+
      #Ethnicity+BMI+LDLC_S+Tab,weight=fgwt, data=fgdata_cv)
#summary(m1_s)
#p-valor 0.4242

#modelo con iteracci n con edad
m1_s <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemS_cat*Age+Sex+
      Comorb+ Ethnicity+BMI+LDLC_S+Tab,weight=fgwt, data=fgdata_cv)
#summary(m1_s)
#p-valor 0.5589

# probamos supuestos del modelo de regresión de cox

test.phdm<-cox.zph(m1_sdm);test.phdm
ggcoxzph(test.phdm)

### test para observaciones influyentes

ggcoxdiagnostics(m1_sdm , type ="dfbeta" , linear.predictions = F,
      ggtheme= theme_bw())

## deviance residuals

ggcoxdiagnostics(m1_sdm , type ="deviance" , linear.predictions = F,
      ggtheme= theme_bw())
```

```

## modelos con remanentes calculados con Martin
#Percentila 75 de remanentes de colesterol
quantile(base_dm$Rem_M)
base_dm$RemM_cat [base_dm$Rem_M<38]<-0;
base_dm$RemM_cat [base_dm$Rem_M>=38]<-1
base_dm$LDLM_s<-scale(base_dm$LDL_M)
table(base_dm$num_comorb, base_dm$Heart_diseases)

## modelo de Fin y Gray para mortalidad cardiovascular ateroscler tica

fgdata_cv <- finegray(Surv(permeth_int, cv_mort) ~ ., data=base_dm,
                      na.action=na.pass)

# modelo simple
m1_mdm_s<- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemM_cat,
                 weight=fgwt, data=fgdata_cv)
summary(m1_mdm_s)

#modelos saturado
m1_mdm <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemM_cat+
               Age+BMI+Sex+LDLM_s+Comorb+Etn+Tab,
               weight=fgwt, data=fgdata_cv)
summary(m1_mdm)
BIC(m1_mdm)
#interaccion con HOMA#
#m1_m <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat*RemM_cat+Age+
              #Sex+num_comorb+Ethnicity+BMI+LDLM_scale+Tab, weight=fgwt,
              #data=fgdata_cv)
#summary(m1_m)
#p-valor 0.6183

```

```
#interaccion edad#
#m1_m <- coxph(Surv(fgstart, fgstop, fgstatus) ~HOMAIR_cat+RemM_cat*Age+
              #Sex+num_comorb+Ethnicity+BMI+LDLM_scale+Tab, weight=fgwt,
              #data=fgdata_cv)
#summary(m1_m)
#p-valor 0.708

# probamos supuestos del modelo de regresión de cox

test.phm<-cox.zph(m1_mdm);test.phm
ggcoxzph(test.phm)
ggcoxzph(test.phm[2])

### test para observaciones influyentes

ggcoxdiagnostics(m1_mdm , type ="dfbeta" , linear.predictions = F,
                 ggtheme= theme_bw())

## deviance residuals

ggcoxdiagnostics(m1_mdm , type ="deviance" , linear.predictions = F,
                 ggtheme= theme_bw())

### modelos de mediación causal para mortalidad aterosclerótica ###

## method=="Aalen" -> output Hazard Difference
## method=="Cox"   -> output Hazard Ratio

mediation_ci1 <- function(lambda.s, lambda.g, covar11, covar12,
                          covar22, alpha.s, var_alpha, G=10^4, method){
```

```
require(mvtnorm)
Omega <- matrix(c(covar11,covar12,covar12,covar22),nrow=2)
IE <- rep(0,G); DE <- rep(0,G); TE <- rep(0,G); Q <- rep(0,G)

set.seed(137)
lambda <- rmvnorm(G, mean = c(lambda.s, lambda.g), sigma = Omega)
alpha <- rnorm(G, mean=alpha.s, sd=sqrt(var_alpha))
DE <- lambda[,1]
IE <- lambda[,2] * alpha
TE <- IE + DE

DE.obs <- lambda.s
IE.obs <- lambda.g * alpha.s
TE.obs <- DE.obs+IE.obs
pval.DE<-2*min(mean((DE-mean(DE))>DE.obs), mean((DE-mean(DE))<DE.obs))
pval.IE<-2*min(mean((IE-mean(IE))>IE.obs), mean((IE-mean(IE))<IE.obs))
pval.TE<-2*min(mean((TE-mean(TE))>TE.obs), mean((TE-mean(TE))<TE.obs))

if (method=="Cox") {DE=exp(DE); IE=exp(IE); TE=exp(TE)}
print("DE:")
print(iffelse(method=="Aalen", DE.obs, exp(DE.obs)))
print(quantile(DE, c(0.025, 0.975)))
print(paste("pval_DE=", pval.DE))
print("IE:")
print(iffelse(method=="Aalen", IE.obs, exp(IE.obs)))
print(quantile(IE, c(0.025, 0.975)))
print(paste("pval_IE=", pval.IE))
print("TE:")
print(iffelse(method=="Aalen", TE.obs, exp(TE.obs)))
print(quantile(TE, c(0.025, 0.975)))
print(paste("pval_TE=", pval.TE))
}
```

```

mediation_ci2 <- function(lambda.s, lambda.m, lambda.g, Sigma.lambda,
                          alpha.s, alpha.m, Sigma.alpha, delta.s, Sigma.delt
                          G=10^4, method){
  require(mvtnorm)
  SY <- rep(0,G); SGY <- rep(0,G); SMY <- rep(0,G); TE <- rep(0,G)
  set.seed(137)
  lambda <- rmvnorm(G, mean = c(lambda.s, lambda.m, lambda.g),
                    sigma = Sigma.lambda)
  alpha <- rmvnorm(G, mean = c(alpha.s, alpha.m),
                  sigma = Sigma.alpha)
  delta <- rnorm(G, mean=delta.s, sd=sqrt(Sigma.delta))
  SY <- lambda[,1]
  SGY <- lambda[,3] * alpha[,1]
  SMY <- (lambda[,2] + lambda[,3]*alpha[,2])*delta
  TE <- SY+SGY+SMY

  SY.obs <- lambda.s
  SGY.obs <- lambda.g * alpha.s
  SMY.obs <- (lambda.m + lambda.g*alpha.m)*delta.s
  TE.obs <- SY.obs+SGY.obs+SMY.obs
  pval.SY<-2*min(mean((SY-mean(SY))>SY.obs), mean((SY-mean(SY))<SY.obs))
  pval.SGY<-2*min(mean((SGY-mean(SGY))>SGY.obs), mean((SGY-mean(SGY))<SGY.ob
  pval.SMY<-2*min(mean((SMY-mean(SMY))>SMY.obs), mean((SMY-mean(SMY))<SMY.ob
  pval.TE<-2*min(mean((TE-mean(TE))>TE.obs), mean((TE-mean(TE))<TE.obs))

  if (method=="Cox") {SY=exp(SY); SGY=exp(SGY); SMY=exp(SMY); TE=exp(TE)}
  print("SY:")
  print(ifelse(method=="Aalen", SY.obs, exp(SY.obs)))
  print(quantile(SY, c(0.025, 0.975)))
  print(paste("pval_SY=", pval.SY))
  print("SGY:")

```

```
print(iffelse(method=="Aalen", SGY.obs, exp(SGY.obs)))
print(quantile(SGY, c(0.025, 0.975)))
print(paste("pval_SGY=", pval.SGY))
print("SMY:")
print(iffelse(method=="Aalen", SMY.obs, exp(SMY.obs)))
print(quantile(SMY, c(0.025, 0.975)))
print(paste("pval_SMY=", pval.SMY))
print("TE:")
print(iffelse(method=="Aalen", TE.obs, exp(TE.obs)))
print(quantile(TE, c(0.025, 0.975)))
print(paste("pval_TE=", pval.TE))
}

## Modelo sujetos sin dm #

# modelo con Sampson
set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
} else if(method=="Cox"){
  lambdas<-m1_s$coef
  lambdas.var<-m1_s$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2], lambdas.var[2,1],
              lambdas.var[1,1],
              m_S$coef[2], vcov(m_S)[2,2], G=10^6, method=method)

IC_inf<-(0.023194/0.120241)*100
IC_Sup<-(0.130998/.777834)*100
Prop_med<-(0.07558/0.411748)*100
```

```
Prop_med;IC_inf;IC_Sup

#### Modelo con Martin
set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
} else if(method=="Cox"){
  lambdas<-m1_m$coef
  lambdas.var<-m1_m$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2], lambdas.var[2,1],
              lambdas.var[1,1],
              m_M$coef[2], vcov(m_M)[2,2], G=10^6, method=method)

IC_infm<-(.020791/.080397)*100
IC_Supm<-(0.115647/0.724770)*100
Prop_medm<-(.067014/0.365573)*100
Prop_medm;IC_infm;IC_Supm

# modelo en dm#
# modelos sampson simple

set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
```

```

} else if(method=="Cox"){
  lambdas<-m1_sdm_s$coef
  lambdas.var<-m1_sdm_s$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2], lambdas.var[2,1],
              lambdas.var[1,1],
              m_Sdm$coef[2], vcov(m_Sdm)[2,2], G=10^6, method=method)

IC_infm<-((.018572)/.186161)*100
IC_Supm<-(0.224864 /0.879103)*100
Prop_medm<-(.11544/0.493259)*100
Prop_medm;IC_infm;IC_Supm

#Sampson

set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
} else if(method=="Cox"){
  lambdas<-m1_sdm$coef
  lambdas.var<-m1_sdm$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2], lambdas.var[2,1],
              lambdas.var[1,1],
              m_Sdm$coef[2], vcov(m_Sdm)[2,2], G=10^6, method=method)

IC_infm<-((1-0.9785013)/.088763)*100
IC_Supm<-(0.1759888/0.741443 )*100
Prop_medm<-(.0717/0.377301)*100
Prop_medm;IC_infm;IC_Supm

```

```
# modelo simple Martin
set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
} else if(method=="Cox"){
  lambdas<-m1_mdm_s$coef
  lambdas.var<-m1_mdm_s$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2], lambdas.var[2,1],
              lambdas.var[1,1],
              m_Mdm$coef[2], vcov(m_Mdm)[2,2], G=10^6, method=method)

IC_infm<-((.018103)/0.200504)*100
IC_Supm<-(0.210672 /0.908964)*100
Prop_medm<-(.108781/0.514316)*100
Prop_medm;IC_infm;IC_Supm

# Martin satturado
set.seed(1234)
method="Cox"
if (method=="Aalen"){
  lambdas<-aalen1$gamma
  lambdas.var<-aalen1$robvar.gamma
} else if(method=="Cox"){
  lambdas<-m1_mdm$coef
  lambdas.var<-m1_mdm$var
}
mediation_ci1(lambdas[2], lambdas[1], lambdas.var[2,2], lambdas.var[2,1],
```

```
      lambdas.var[1,1],  
      m_Mdm$coef[2], vcov(m_Mdm)[2,2], G=10^6, method=method)  
  
IC_infm<-((1-0.9749823)/0.080893)*100  
IC_Supm<-(0.1622153 /0.752082)*100  
Prop_medm<-(0.063621/0.37663)*100  
Prop_medm;IC_infm;IC_Supm
```