GRAFOS ACÍCLICOS DIRIGIDOS

Teoría y Práctica en Epidemiología Laboral

Galina Diaz Barrientos
María Victoria Lizarbe Castro
Katherine Gisselle Noriega Herrera
César Jesús Eras Lévano



NOVO MUNDO

Galina Diaz Barrientos
Perfil: Médico cirujano con maestría en Salud Pública, maestría en Salud Ocupacional y Ambiental, docente universitario y conferencista.
ORCID: https://orcid.org/0000-0002-8502-5842
Afiliación: Universidad Norbert Wiener.

María Victoria Lizarbe Castro
Perfil: Licenciada en Enfermería, Maestra en Salud Pública, Doctora en Salud Pública y Especialista en Epidemiología de Campo.
ORCID: https://orcid.org/0000-0002-4259-3399
Afiliación: Universidad Privada San Juan Bautista.

Katherine Gisselle Noriega Herrera
Perfil: Médico cirujano, Magíster en Salud Ocupacional y Ambiental, con formación en neurociencia aplicada.
ORCID: https://orcid.org/0009-0007-6024-337X
Afiliación: Fuente 360.

César Jesús Eras Lévano
Perfil: Médico cirujano, Magíster en Salud Ocupacional y docente universitario.
ORCID: https://orcid.org/0000-0002-5433-4044
Afiliación: Universidad Privada San Juan Bautista.

GRAFOS ACÍCLICOS DIRIGIDOS

Teoría y Práctica en Epidemiología Laboral

Galina Diaz Barrientos

María Victoria Lizarbe Castro

Katherine Gisselle Noriega Herrera

César Jesús Eras Lévano

NOVO MUNDO

Todas nuestras publicaciones son sometidas a revisión doble-ciego de pares externos (*Peer Review Double Blinded*).

Esta publicación cuenta con licencia Creative Commons Reconocimiento - NoComercial - SinObraDerivada 3.0 Unported License.



ISBN: 978-9942-694-23-2

© Galina Diaz Barrientos
© María Victoria Lizarbe Castro
© Katherine Gisselle Noriega Herrera
© César Jesús Eras Lévano

2026
DOI: https://doi.org/10.5281/zenodo.21478086
© Editorial Novo Mundo
Guayaquil, Ecuador.
www.editorialnovomundo.com

Queda prohibida la reproducción bajo cualquier modalidad de toda o una parte de esta obra sin autorización expresa del titular de los derechos.

Diseño de carátula y composición: Editorial Novo Mundo
Edición electrónica: Editorial Novo Mundo

Editado en Ecuador/ *Published in Ecuador*

GRAFOS ACÍCLICOS DIRIGIDOS

Teoría y Práctica en Epidemiología Laboral

Galina Diaz Barrientos

María Victoria Lizarbe Castro

Katherine Gisselle Noriega Herrera

César Jesús Eras Lévano

NOVO MUNDO

ÍNDICE

PREFACIO **13**

INTRODUCCIÓN GENERAL **14**

CAPÍTULO I.

INTRODUCCIÓN A LOS GRAFOS ACÍCLICOS DIRIGIDOS Y SU RELEVANCIA EN LA EPIDEMIOLOGÍA LABORAL **17**

- Ejemplo 1.1: DAG Básico para Exposición a Sílice y EPOC(Enfermedad Pulmonar Obstructiva Crónica) 18
- Ejemplo 1.2: Comparación de Estrategias de Ajuste (Convencional vs DAG) 22

CAPÍTULO II.

FUNDAMENTOS TEÓRICOS: TEORÍA CAUSAL Y REPRESENTACIÓN GRÁFICA **28**

2.1. Orígenes y Desarrollo de los DAG en Inferencia Causal 28

2.2. Componentes Estructurales de un DAG 29

2.3. Supuestos Fundamentales 30

- Ejemplo 2.1: Verificación de d-separación y bloqueo de caminos 30
- Ejemplo 2.2: Efecto total vs efecto directo 34

2.4. Diferenciación entre Asociación y Causalidad 36

CAPÍTULO III.

IDENTIFICACIÓN Y CONTROL DE SESGOS MEDIANTE DAG: CONFUSIÓN, COLISIONADORES Y SELECCIÓN 38

3.1. Confusión y el Conjunto de Ajuste Mínimo 38

3.2. Sesgo de Colisión (Collider Bias) 39

3.3. Sesgo de Selección y Muestreo 39

3.4. Reglas Prácticas para el Uso de DAG en Control de Sesgos 40

- Ejemplo 3.1: Sesgo por colisión al ajustar por baja laboral 40
- Ejemplo 3.2: Healthy Worker Survivor Effect (sesgo de selección) 44

CAPÍTULO IV.

DAG EN LA EPIDEMIOLOGÍA LABORAL: CONTEXTO, EXPOSICIONES Y RESULTADOS OCUPACIONALES 50

4.1. Especificidades del Entorno Ocupacional 50

4.2. Ejemplos Estructurados en Exposiciones Clásicas 51

4.2.1. Sílice cristalina y EPOC 51

4.2.2. Trabajo por turnos y enfermedad cardiovascular 51

4.2.3. Estrés psicosocial y trastornos musculoesqueléticos 51

4.3. Integración con Vigilancia Ocupacional.....52

- Ejemplo 4.1: Exposición mixta ergonómica y psicosocial.....52
- Ejemplo 4.2: Alteración del ritmo circadiano y fatiga laboral 57

CAPÍTULO V.

DISEÑO DE ESTUDIOS Y ANÁLISIS CAUSAL BASADO EN DAG EN ENTORNOS OCUPACIONALES.....63

5.1. Cohortes Ocupacionales y Registros de Exposición.....63

5.2. Estudios Transversales y de Prevalencia.....64

5.3. Metaanálisis y Síntesis de Evidencia 64

5.4. Métodos Estadísticos Compatibles con DAG 64

- Ejemplo 5.1: Caso-control emparejado por edad y sexo 65

CAPÍTULO VI.

IMPLEMENTACIÓN PRÁCTICA: SOFTWARE, VALIDACIÓN Y COMUNICACIÓN DE RESULTADOS 70

6.1. Herramientas de Software Especializado.....70

6.2. Proceso de Validación y Sensibilidad.....71

6.3. Comunicación a Stakeholders Ocupacionales.....71

- Ejemplo 6.1. De Flujo de Trabajo en Rstudio 72
- Ejemplo 6.2: Flujo completo en R (DAG → fórmula → validación 73

CAPÍTULO VII.

LIMITACIONES, CRÍTICAS Y CONSIDERACIONES ÉTICAS EN EL USO DE DAG 80

7.1. Limitaciones Metodológicas 80

7.2. Críticas Académicas 81

7.3. Consideraciones Éticas en Salud Ocupacional 81

- Ejemplo 7.1: Confusor no medido y análisis de sensibilidad 82
- Ejemplo 7.2: Variables proxy y equidad en salud laboral 86

7.4 Integración del Modelo Demandas-Recursos Laborales (JD-R) en Grafos Causales 92

7.4.1 El Modelo JD-R como Marco Teórico Causal 92

7.4.2 Representación del JD-R en un DAG 92

7.4.3 El Desafío de la Interacción (Efecto Buffer) en DAG 93

7.4.4 Ejemplo Práctico: Personal Sanitario y Burnout 94

7.4.5 Limitaciones Causales del JD-R 96

CAPÍTULO VIII.

PERSPECTIVAS FUTURAS Y REFLEXIONES: INTEGRACIÓN DE DAG EN LA SALUD OCUPACIONAL MODERNA 97

8.1. Tendencias Emergentes 97

8.2. Recomendaciones para la Práctica 98

- Ejemplo 8.1: DAG dinámico con medidas repetidas 98
- Ejemplo 8.2: Fusión causal para revisiones sistemáticas 103

8.3 Comparación Metodológica entre DAG y Modelos de Ecuaciones Estructurales (SEM 109

8.3.1 Contexto: El Dominio del SEM en Psicología de la Salud Ocupacional 109

8.3.2 Diferencias Fundamentales entre DAG y SEM 110

8.3.3 Fortalezas de los DAG sobre SEM en Epidemiología Ocupacional 110

CAPÍTULO IX.

INFERENCIA CAUSAL EN LA EVALUACIÓN DE TERAPIAS DE LARGA LATENCIA: UNA FORMALIZACIÓN TEÓRICA DEL SESGO DE TIEMPO DE INMORTALIDAD MEDIANTE GRAFOS ACÍCLICOS DIRIGIDOS 112

9.1. Introducción 112

9.2. Aspectos metodológicos 114

9.3. Resultados 117

9.3.1. Ontología del tiempo en el DAG 117

9.3.2. Mecanismos de selección dinámica 118

9.3.3. Soluciones estructurales 119

9.3.4. Implicaciones en la toma de decisiones 120

9.3.5. Sesgo de "tiempo inmortal" en la evaluación de terapias de larga latencia 120

9.4. Discusión 121

CONCLUSIÓN 126

REFERENCIAS BIBLIOGRÁFICAS 127

GRÁFICOS

Gráfico 1.
DAG Básico- Exposición a Silice y EPOC 21

Gráfico 2.
DAG/ Estrategias de Ajuste (Convencional vs DAG 26

Gráfico 3.
DAG/Exposición a Ruido Industrial sobre Pérdida Auditiva Neurosensoria.....32

Gráfico 4.
DAG / Con Nodos Coloreados por Rol Causal 33

Gráfico 5.
DAG/Efecto total vs efecto directo 35

Gráfico 6.
DAG/ Sesgo por colisión al ajustar por baja laboral 44

Gráfico 7.
DAG/ Healthy Worker Survivor Effect (sesgo de selección) 48

Gráfico 8.
DAG/ Exposición mixta ergonómica y psicosocial 56

Gráfico 9.
DAG/ Alteración del ritmo circadiano y fatiga laboral.....61

Gráfico 10.
DAG/Caso-control emparejado por edad y sexo.....68

Gráfico 11.
DAG/De Flujo de Trabajo en Rstudio.....73

Gráfico 12.
DAG/Estudio transversal nacional sobre
vibraciones mano-brazo y síndrome de Raynaud.....77

Gráfico 13.
DAG/ Confusor no medido y análisis de sensibilidad.....85

Gráfico 14.
DAG/Variables proxy y equidad en salud laboral.....90

Gráfico 15.
DAG/Modelo Causal JD-R.....95

Gráfico 16.
DAG/Dinámico con medidas repetidas.....102

Gráfico 17.
DAG/ Fusión causal para revisiones sistemáticas.....108

TABLAS

Tabla 1.
Elementos de un Modelo Causal JD-R 95

Tabla 2.
Diferencias Fundamentales entre DAG y SEM.....110

FIGURAS

Figura 1.
Componentes fundamentales del DAG 29

PREFACIO

La epidemiología ocupacional se enfrenta a desafíos únicos en la identificación de relaciones causales entre exposiciones laborales y efectos adversos para la salud. A diferencia de los ensayos clínicos aleatorizados, los estudios observacionales en el lugar de trabajo están plagados de sesgos complejos, como el "Efecto del Trabajador Sano", la confusión por factores socioeconómicos y la medición errónea de la exposición. Durante décadas, el control de estas variables se basó en criterios estadísticos tradicionales que, a menudo, introducían más sesgo del que eliminaban.

La introducción de los Grafos Acíclicos Dirigidos (DAG, por sus siglas en inglés) ha revolucionado la forma en que conceptualizamos la causalidad. Este libro no es solo un manual técnico; es una guía práctica para el médico laboral y el epidemiólogo que busca rigor científico. A lo largo de ocho capítulos, desglosaremos la teoría de grafos y la aplicaremos a escenarios reales: desde la silicosis hasta los trastornos musculoesqueléticos y la salud mental laboral.

Esta investigación busca proporcionar una herramienta para el diseño de estudios más robustos, el análisis de datos más preciso y, fundamentalmente, la toma de decisiones en prevención de riesgos laborales basada en evidencia causal sólida. Por tal razon, el objetivo del estudio presentado es: Desarrollar un manual y guía académico-práctica con base en evidencia científica y metodológica sobre el uso de los DAG en la epidemiología laboral.

Los autores

INTRODUCCIÓN GENERAL

Gutiérrez-Strauss y el Ministerio de la Protección Social de Colombia (2011) reseñan que en diversos sectores económicos y laborales se presentan escenarios de riesgos laborales que vulneran a la población trabajadora con importantes cargas de enfermedades. Estos riesgos y padecimientos se concentran tanto en trabajadores que laboran como agricultores, operarios de maquinaria, técnicos ejecutores u operantes y profesionales de la salud, reflejando globalmente la desigualdad estructural ocupacional expuesta.

Para la Organización Mundial de la Salud [OMS] (2022), entre el 30% y el 50% de la población mundial que labora está expuesta a factores de riesgo físicos, químicos, biológicos, psicológicos, sociales y ergonómicos que perjudican la integridad física y la capacidad productiva de los trabajadores; siendo las patologías crónicas más comunes la enfermedad pulmonar obstructiva crónica (EPOC), las dorsalgias (dolor de espalda), el asma, la depresión y la ansiedad.

La Universidad Católica San Pablo [UCSP] (2026) refiere que son millones de trabajadores los que mueren cada año por afecciones no transmisibles provocadas por las dinámicas laborales. Las más riesgosas para la morbilidad laboral son: la exposición a jornadas laborales extenuantes, el aire contaminado en los lugares de trabajo, las deficientes condiciones ergonómicas y las presiones psicosociales.

El Instituto Nacional de Seguridad e Higiene en el Trabajo de España [INSST] (2026) refiere que la epidemiología laboral —como disciplina que estudia los efectos por exposición laboral, así como las frecuencias y distribución de los padecimientos y lesiones ocupacionales— ha enfocado su pensamiento y metodología de observación en ahondar en las verdaderas causas de los padecimientos laborales, trascendiendo toda perspectiva tradicional biomédica. La labor de esta disciplina científica ha sido desconceptualizar la salud como la simple inexistencia de la enfermedad o disfuncionalidad, para comprenderla como un estado del ser humano que abarca autonomía, bienestar y dignidad; por ello, defender y conservar la salud laboral se traduce en un signo de calidad de vida individual que resulta en evolución personal, bienestar familiar y un equilibrio socioambiental de las poblaciones (INSST, 2026).

A partir de los hallazgos y estudios de la epidemiología laboral con respecto a las diversas enfermedades ocupacionales, la Organización Internacional del Trabajo [OIT] (2010), en aras de resaltar la importante labor hecha por esta disciplina en el área, recoge en un epígrafe tanto la identificación como el reconocimiento de las enfermedades profesionales. Esto constituye a esta rama de estudio como indispensable para la valoración relacional entre condiciones ocupacionales, hipótesis causales y la salud ocupacional, con el fin de que las medidas preventivas y la toma de decisiones sean más eficaces y efectivas en los entornos laborales (INSST, 2026).

En la actualidad, una de las herramientas metodológicas utilizadas en los estudios observacionales de la epidemiología laboral son los Grafos Acíclicos Dirigidos (DAG), ya que permiten identificar, dentro de un conjunto de variables, un mínimo pero acertadas de ellas, que al poder controlarse, permiten obtener el verdadero efecto causal de la exposición ocupacional sobre la salud de los trabajadores; ayudando a los investigadores al realizar una regresión, saturarse o confundirse considerando todas las variables disponibles. Los DAG pasaron de ser fundamentos teóricos en los campos de estudio de la estadística e informática a convertirse en eficaces herramientas, tanto para el diseño de estudios observacionales como para dis-

minuir los sesgos y aseverar la validez de la inferencia causal (Fox et al., 2024). Por tal razón, el objetivo general de esta investigación es: desarrollar un manual y guía académico-práctica con base en evidencia científica y metodológica sobre el uso de los Grafos Acíclicos Dirigidos (DAG) en la epidemiología laboral.

Para ello, la investigación desarrolla una estructura de contenido de 9 capítulos que abarcan desde los fundamentos teóricos hasta la implementación práctica, pasando por la identificación de sesgos, el diseño de estudios, la validación mediante software especializado, las limitaciones éticas y metodológicas, las perspectivas de integración en la vigilancia y prevención ocupacional, y un último capítulo de investigación donde se formaliza teóricamente, en el campo de la inferencia causal, el sesgo de tiempo de inmortalidad mediante DAG evaluando terapias de larga latencia.

CAPÍTULO I

INTRODUCCIÓN A LOS GRAFOS ACÍCLICOS DIRIGIDOS Y SU RELEVANCIA EN LA EPIDEMIOLOGÍA LABORAL

La epidemiología laboral se enfrenta a desafíos metodológicos únicos: exposiciones complejas y multifactoriales, poblaciones dinámicas, sesgos de selección derivados del empleo activo, y la dificultad de establecer relaciones causales en entornos donde la asignación a la exposición no es aleatoria. Históricamente, el control de factores de confusión se ha basado en criterios estadísticos convencionales (p. ej., cambio en el estimador >10%, significancia estadística), enfoques que han demostrado ser insuficientes o incluso contraproducentes cuando la estructura causal subyacente no se especifica explícitamente (Greenland et al., 1999; Schisterman et al., 2009).

Los Grafos Acíclicos Dirigidos (DAG, por sus siglas en inglés: *Directed Acyclic Graphs*) emergen como una herramienta formal de inferencia causal que permite representar supuestos teóricos sobre las relaciones entre variables, identificar conjuntos mínimos de covariables que deben ajustarse para eliminar el sesgo de confusión, y detectar estructuras que introducen sesgo cuando se ajustan incorrectamente (p. ej., colisionadores o *colliders*)

(Pearl, 1995; Hernán & Robins, 2020). Su adopción en epidemiología ha transformado la manera en que se diseñan, analizan e interpretan los estudios observacionales, particularmente en contextos laborales donde la validez interna depende críticamente de decisiones de ajuste adecuadas.

En la salud ocupacional, el uso de DAG facilita la traducción de conocimiento experto (toxicología, ergonomía, psicología del trabajo) en estructuras gráficas validables, permitiendo a los investigadores distinguir entre variables que actúan como confusores, mediadores, colisionadores o descendientes de la exposición. Esta distinción es crucial para evitar sesgos como el de colisión, frecuentemente introducido al ajustar por factores que son consecuencia tanto de la exposición como del resultado (Cole et al., 2010; VanderWeele, 2019).

Ejemplo 1.1: DAG Básico para Exposición a Sílice y EPOC(Enfermedad Pulmonar Obstructiva Crónica)

- **Contexto:** Estudio de cohortes en trabajadores de construcción expuestos a polvo de sílice cristalina. Se busca estimar el efecto causal total sobre EPOC (ver Gráfico 1)
- **Código en Estudio:**

```
> # EJEMPLO 1.1: DAG BÁSICO PARA EXPOSICIÓN A SÍLICE Y EPOC
> # Capítulo 1: Introducción a los DAG en Epidemiología Laboral
> library(dagitty)
> library(ggdag)
> library(ggplot2)
 # ------------------------------------------------------------------------
> # 1. DEFINICIÓN DEL DAG
> # ------------------------------------------------------------------------
> dag_silica <- dagitty('dag {
+   silica [exposure]
+   copd [outcome]
```

```
+   age [adjusted]
+   smoking [adjusted]
+   ses [adjusted]
+   resp_protection [selected]
+   age -> silica
+   age -> copd
+   smoking -> silica
+   smoking -> copd
+   ses -> smoking
+   ses -> silica
+   silica -> copd
+   silica -> resp_protection
+   copd -> resp_protection}')
> # ------------------------------------------------------------------------
> # 2. VERIFICACIÓN METODOLÓGICA
> # ------------------------------------------------------------------------
> isAcyclic(dag_silica)
[1] TRUE
> adjustmentSets(dag_silica, exposure = "silica", outcome = "copd", effect
= "total")
> # ------------------------------------------------------------------------
> # 3. VISUALIZACIÓN
> # ------------------------------------------------------------------------
> dag_tidy <- tidy_dagitty(dag_silica)
> # Definir etiquetas en español (versión robusta)
> etiquetas <- c(
+     silica = "Sílice\nCristalina",
+     copd = "EPOC",
+     age = "Edad",
+     smoking = "Tabaquismo",
+     ses = "Nivel\nSocioeconómico",
+     resp_protection = "Uso de\nEPP\nRespiratorio")
> # Colores por tipo de variable (paleta accesible)
> colores <- c(
```

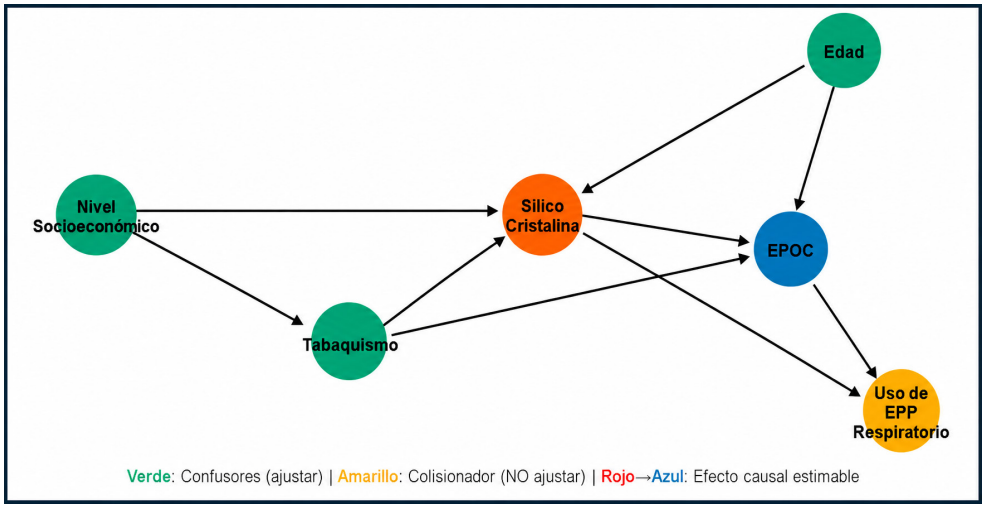
```
+     silica = "#D55E00",         # Exposición: Rojo-naranja
+     copd = "#0072B2",           # Outcome: Azul
+     age = "#009E73",            # Confusor: Verde
+     smoking = "#009E73",        # Confusor: Verde
+     ses = "#009E73",            # Confusor: Verde
+     resp_protection = "#E69F00"  # Colisionador: Amarillo)
> # Crear el gráfico base usando la API nativa de ggdag (más estable)
> p <- ggdag(dag_tidy,
+            node_size = 16,
+            text_size = 3.5,
+            text_col = "white",
+            edge_type = "link") +  # Tipo de arista simple y limpio
+
+     # Mapeo manual de colores
+     geom_dag_point(aes(color = name), size = 18) +
+     scale_color_manual(values = colores) +
+      # Etiquetas personalizadas usando data subyacente (evita errores de
longitud)
+     geom_text(data = dag_tidy$data %>% filter(!is.na(name)),
+               aes(x = x, y = y, label = etiquetas[name]),
+               color = "Black",
+               fontface = "bold",
+               size = 3.3,
+               lineheight = 0.85) +
+     # Tema académico
+     theme_dag(base_size = 12) +
+     labs(
+       title = "Figura 1.1: DAG Básico - Exposición a Sílice y EPOC",
+       subtitle = "Estudio de cohortes en trabajadores de construcción",
+         caption = "Verde: Confusores (ajustar) | Amarillo: Colisionador
(NO ajustar) | Rojo→Azul: Efecto causal estimable") +
+     theme(
+       plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
+      plot.subtitle = element_text(hjust = 0.5, color = "gray30"),
```

```
+       plot.caption = element_text(hjust = 0.5, size = 10, color = "gray40"),
+       legend.position = "none")
> print(p)
```

Código en R generado por el autor.

Gráfico 1

DAG Básico- Exposición a Silice y EPOC



Nota. Elaboración propia. Estudios de cohortes en trabajadores de construcción

Explicación del DAG:

En el centro está lo que queremos saber: ¿El polvo de sílice cristalina (el naranja) causa EPOC (el azul)?

Esa flecha directa es la ruta verdadera que queremos medir. Es como preguntar: "Si controlo la exposición a sílice, ¿en cuánto baja el riesgo de enfermedad pulmonar?"

Además, hay tres "caminos alternativos" (líneas verdes) que crean confusión estadística (confusores):

- **Edad**: Los trabajadores mayores llevan más años expuestos a sílice (más tiempo para acumular polvo), pero también sus pulmones están más "gastados" por el tiempo. Si no controlamos la edad, podríamos pensar que la sílice causa más daño del que realmente causa, o menos.
- **Tabaquismo**: Fumar daña los pulmones (va directo a EPOC), pero también suele ser más común en ciertos oficios de la construcción. Es una "tercera variable" que ensucia la relación.
- **Nivel Socioeconómico**: Determina qué tipo de trabajo realizas (más o menos exposición a polvo) y también está relacionado con si fumas o no.

  Ahora viene la parte más contra-intuitiva: Uso de EPP Respiratorio (Círculo amarillo).

  Observa las flechas que entran a esa variable:

  - Viene de la Sílice (cuando hay mucho polvo, te pones la mascarilla)
  - Viene de la EPOC (cuando ya estás enfermo, el doctor te obliga a usar protección)

Esto crea un colisionador (como un choque de trenes). Si tú decides "ajustar" o controlar por el uso de mascarilla (por ejemplo, comparando solo a los que usan mascarilla vs. los que no), ¡Se apertura un camino falso entre la sílice y la EPOC!

Ejemplo 1.2: Comparación de Estrategias de Ajuste (Convencional vs DAG)

- **Contexto:** Re-análisis de un estudio transversal sobre trabajo nocturno y síndrome metabólico, donde originalmente se ajustó por IMC, antigüedad y participación en programas de bienestar (ver Gráfico 2).

- **Código para Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_shift <- dagitty('dag {
  shift_work [exposure]
  metabolic_syndrome [outcome]
  age [adjusted]
  sex [adjusted]
  education [adjusted]
  lifestyle_score [adjusted]
  bmi [mediator]
  wellness_program [collider]
  job_tenure [adjusted]
  age -> shift_work
  age -> metabolic_syndrome
  age -> job_tenure
  sex -> shift_work
  sex -> metabolic_syndrome
  education -> shift_work
  education -> lifestyle_score
  education -> metabolic_syndrome
  lifestyle_score -> metabolic_syndrome
  lifestyle_score -> bmi
  shift_work -> bmi
  bmi -> metabolic_syndrome
  shift_work -> metabolic_syndrome
  shift_work -> wellness_program
  metabolic_syndrome -> wellness_program
  job_tenure -> shift_work}')
dag_tidy <- tidy_dagitty(dag_shift)
dag_tidy$data <- dag_tidy$data %>%
```

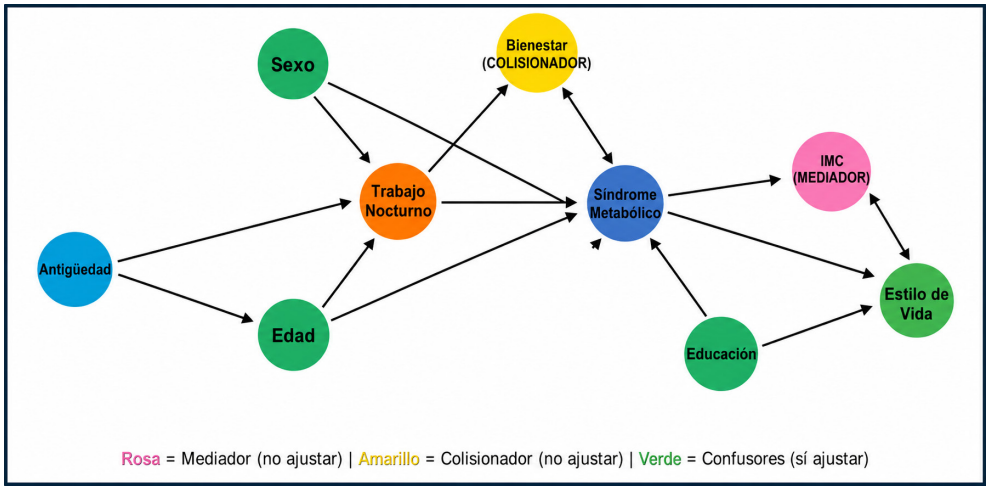
```
mutate(
    x = x * 2.2,
    y = y * 1.8,
xend = xend * 2.2,
    yend = yend * 1.8)
etiquetas <- c(
  shift_work = “Trabajo\nNocturno”,
  metabolic_syndrome = “Síndrome\nMetabólico”,
  bmi = “IMC\n(MEDIADOR)”,
  wellness_program = “Programa\nBienestar\n(COLISIONADOR)”,
  age = “Edad”,
sex = “Sexo”,
  education = “Educación”,
  lifestyle_score = “Estilo de\nVida”,
  job_tenure = “Antigüedad”)
colores <- c(
shift_work = “#D55E00”,
  metabolic_syndrome = “#0072B2”,
  bmi = “#CC79A7”,
  wellness_program = “#F0E442”,
  age = “#009E73”,
  sex = “#009E73”,
  education = “#009E73”,
  lifestyle_score = “#009E73”,
  job_tenure = “#56B4E9”)
p <- ggplot(dag_tidy$data, aes(x = x, y = y, xend = xend, yend = yend)) +
  geom_dag_edges(edge_color = “gray40”, edge_width = 0.8) +
  geom_dag_point(aes(color = name), size = 16) +
  scale_color_manual(values = colores) +
  geom_text(
    data = filter(dag_tidy$data, !is.na(name)),
    aes(label = etiquetas[name]),
    color = “black”,
    fontface = “bold”,
```

```
    size = 3.4,
    lineheight = 0.85) +
  coord_cartesian(
    xlim = c(min(dag_tidy$data$x, na.rm = TRUE) - 0.8,
             max(dag_tidy$data$x, na.rm = TRUE) + 0.8),
    ylim = c(min(dag_tidy$data$y, na.rm = TRUE) - 0.8,
             max(dag_tidy$data$y, na.rm = TRUE) + 0.8),
clip = "off") +
  labs(
    title = "Figura 1.2: Estrategia Convencional vs. DAG",
    subtitle = "Errores: Ajustar por IMC (mediador) y Programa Bienestar
(colisionador)",
    caption = "Rosa = Mediador (no ajustar) | Amarillo = Colisionador (no
ajustar) | Verde = Confusores (sí ajustar)") +
  theme_dag() +
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5, margin =
margin(b = 15)),
     plot.subtitle = element_text(hjust = 0.5, color = "gray30", size = 11,
margin = margin(b = 20)),
     plot.caption = element_text(hjust = 0.5, size = 10, color = "gray40",
margin = margin(t = 25)),
    plot.margin = margin(40, 40, 40, 40),
    legend.position = "none")
print(p)
ggsave("figura_1_2_dag_shift.png", plot = p, width = 14, height = 10, dpi
= 300, bg = "white")
ggsave("figura_1_2_dag_shift.pdf", plot = p, width = 14, height = 10)
```

Código en Rstudio generado por el autor

Gráfico 2

DAG/ Estrategias de Ajuste (Convencional vs DAG)



Nota: Elaboración propia. En el gráfico representa estudio transversal sobre trabajo nocturno y síndrome metabólico, donde originalmente se ajustó por IMC (mediador), antigüedad y participación en programas de bienestar (colisionador).

A partir del gráfico 2, se infiere que:

El Error del IMC (Mediador):

En estudios sobre trabajo nocturno, el Índice de Masa Corporal suele tratarse como un "confusor" que hay que "controlar". Esto es conceptualmente incorrecto: el trabajo nocturno altera el metabolismo (vía disrupción circadiana y cortisol), lo que causa aumento de peso, y este a su vez contribuye al síndrome metabólico. Por tanto, el IMC es un mediador en la ruta causal, no una causa previa común. Ajustar estadísticamente por IMC equivale a preguntar: "¿Cuál es el efecto del trabajo nocturno independientemente de que engorde?" Esto responde una pregunta clínica diferente (efecto directo) y típicamente subestima el efecto total de la exposición ocupacional.

El Error del Programa de Bienestar (Colisionador);

"Participación en programas de bienestar"parece una variable de control inocua, pero es un colisionador clásico: depende tanto de la exposición (las empresas con turnos nocturnos suelen implementar estos programas por ley o negociación sindical) como del resultado (los trabajadores con síndrome metabólico son derivados a estos programas médicamente). Ajustar por esta variable introduce una asociación espuria entre trabajo nocturno y salud, ya que seleccionamos una subpoblación no representativa (los que participan son sistemáticamente diferentes).

Lección Metodológica:

La estrategia convencional original probablemente concluyó que "el trabajo nocturno no está asociado con síndrome metabólico" o que "la asociación se atenúa tras ajustar por IMC", cuando en realidad esa atenuación era un artefacto metodológico del sobresjuste (overadjustment bias). El enfoque DAG revela que el efecto total solo se estima correctamente cuando se bloquean los caminos de confusión (edad, educación, estilo de vida) sin bloquear los caminos mediadores ni abrir caminos de colisión.

Iremos explorando este contexto y los términos, de ambos ejemplos, en los subsecuentes capítulos.

CAPÍTULO II

FUNDAMENTOS TEÓRICOS: TEORÍA CAUSAL Y REPRESENTACIÓN GRÁFICA

2.1. Orígenes y Desarrollo de los DAG en Inferencia Causal

Los DAG tienen su origen en la teoría de probabilidades condicionales y la representación gráfica de dependencias, desarrollada inicialmente por Judea Pearl en el contexto de la inteligencia artificial y la ciencia de la computación (Pearl, 1988, 1995). Su incorporación a la epidemiología se consolidó a finales de la década de 1990, cuando investigadores como Sander Greenland y James Robins demostraron que los diagramas causales podían formalizar supuestos sobre confusión y selección, y proporcionar reglas algorítmicas para identificar conjuntos de ajuste válidos (Greenland et al., 1999; Robins, 1987).

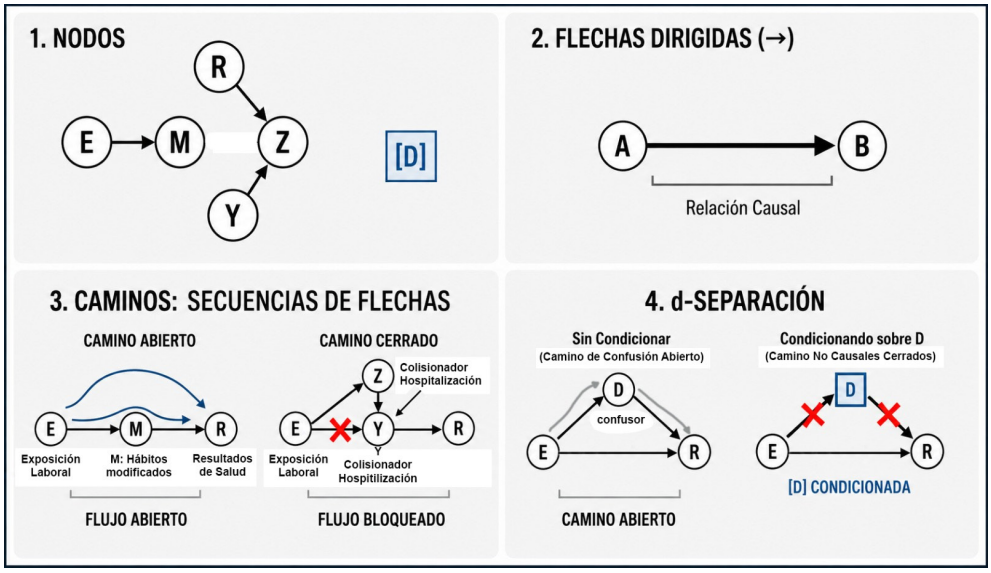
Un DAG es un grafo compuesto por nodos (variables) y aristas dirigidas (flechas) que representan relaciones causales hipotéticas. La condición de aciclicidad implica que no existen ciclos dirigidos; es decir, una variable no puede ser causa directa o indirecta de sí misma en el mismo nivel temporal. Esta restricción es coherente con la noción epidemiológica de precedencia temporal causal (Hernán & Robins, 2020).

2.2. Componentes Estructurales de un DAG (Figura 1).

- **Nodos:** Representan variables observadas o no observadas (p. ej., exposición laboral, resultado de salud, factores demográficos, hábitos).
- **Flechas dirigidas (→):** Indican una relación causal directa hipotética. La dirección implica temporalidad y mecanismo.
- **Caminos dirigidos:** Secuencias de flechas que conectan exposición y resultado. Un camino es *abierto* si no está bloqueado por colisionadores o variables condicionadas; es *cerrado* si el flujo de asociación está interrumpido.
- **d-separación (separación direccional):** Criterio formal que determina si un conjunto de variables cierra todos los caminos no causales entre exposición y resultado, permitiendo estimar el efecto causal total o directo (Pearl, 1995; VanderWeele, 2015)(ver Figura 1).

Figura 1

Componentes fundamentales del DAG



Nota: Elaboración propia a partir de Judea Pearl (1995) y Tyler VanderWeele (2015)

2.3. Supuestos Fundamentales

El uso de DAG en epidemiología laboral reposa sobre supuestos explícitos:

1) **Correcta especificación causal:** El DAG refleja fielmente las relaciones causales conocidas o hipotetizadas. Errores de omisión o dirección invalidan las inferencias (Greenland et al., 1999).

2) **Suficiencia causal (Markov condition):** Dado el conjunto de padres directos de una variable, esta es independiente de todos sus no descendientes.

3) **Fidelidad (faithfulness):** Las independencias estadísticas observadas reflejan independencias estructurales en el DAG, no cancelaciones accidentales de efectos (Spirtes et al., 2000).

4) **No hay efectos no medidos críticos no representados:** Si existen variables no medidas que abren caminos de confusión, el DAG debe incluirlas como nodos latentes o al menos reconocer su impacto potencial (VanderWeele & Shpitser, 2011).

En el contexto laboral, estos supuestos exigen la integración de conocimiento multidisciplinario. Por ejemplo, un DAG que evalúa el efecto de la exposición a sílice cristalina sobre la enfermedad pulmonar obstructiva crónica (EPOC) debe incluir el tabaquismo como confusor clásico, pero también considerar variables como el nivel socioeconómico, el tipo de empleo, la antigüedad laboral y el uso de equipo de protección personal, cuya posición causal determina si su ajuste es necesario, perjudicial o irrelevante (Steenland & Ward, 2014; Checkoway et al., 2004).

Ejemplo 2.1: Verificación de d-separación y bloqueo de caminos

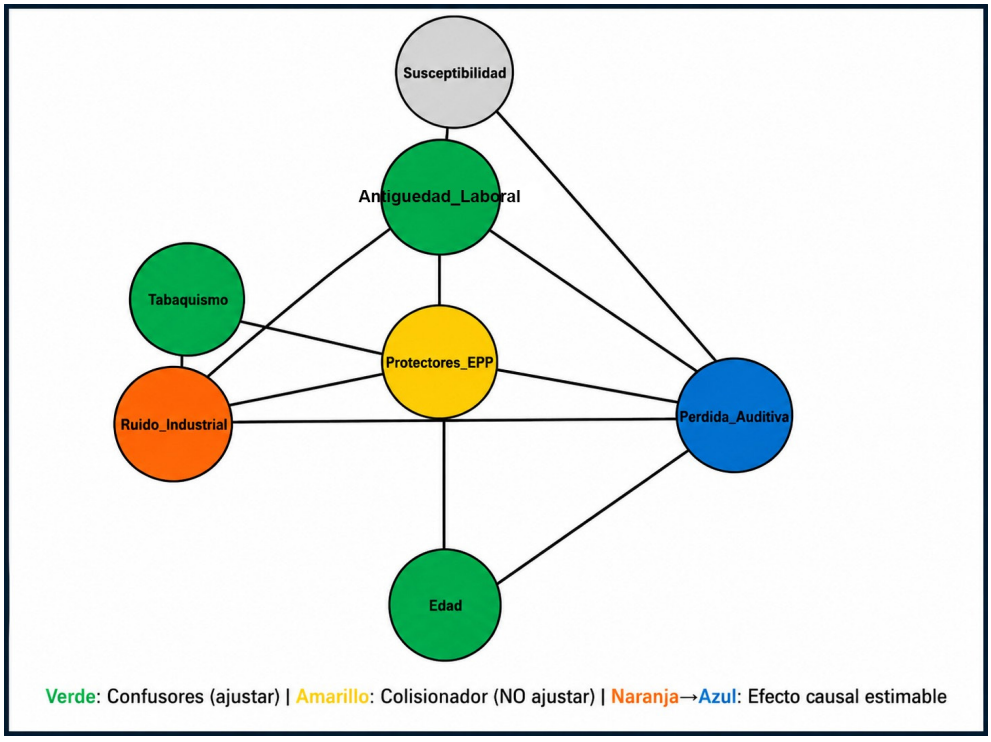
- **Contexto:** Evaluación del efecto de exposición a ruido industrial sobre pérdida auditiva neurosensorial (ver Gráfico 3).
- **Código para Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
#    "Historial" es una abreviatura de "HistorialOtologico"
dag_ruido <- dagitty("dag {
  Ruido [exposure]
  PerdidAuditiva [outcome]
  Edad Sexo UsoEPP Historial
  Ruido -> PerdidAuditiva
  Edad -> Ruido; Edad -> PerdidAuditiva
  Sexo -> Ruido; Sexo -> PerdidAuditiva
  UsoEPP -> PerdidAuditiva
  Ruido -> UsoEPP
  Historial -> PerdidAuditiva}")
#    Gráfico simple con todos los nodos del mismo color
ggdag(dag_ruido, layout = "fr", text = FALSE, node_size = 16) +
  geom_dag_text(color = "black", size = 4) +
theme_dag() +
  labs(title = "DAG: Ruido ocupacional → Pérdida auditiva",
       subtitle = "Relaciones causales postuladas")
#    Nodos coloreados por rol causal (exposición, resultado, confusor, mediador,
etc.)
ggdag_status(dag_ruido, layout = "fr", text = FALSE, node_size = 16) +
  geom_dag_text(color = "black", size = 4, fontface = "plain") +
theme_dag() +
  labs(title = "DAG: Ruido ocupacional como exposición y pérdida auditi-
va como resultado",
       subtitle = "Nodos coloreados según su rol causal",
       caption = "Nota: 'Historial' se refiere a historial otológico previo.")
ggsave("dag_ruido_auditiva.png", width = 8, height = 6, dpi = 300)
```

Código para Rstudio generado por el autor.

Gráfico 3

DAG/Exposición a Ruido Industrial sobre Pérdida Auditiva Neurosensoria



Nota. Elaboración propia. Representa la verificación de d-separación y bloqueo de caminos

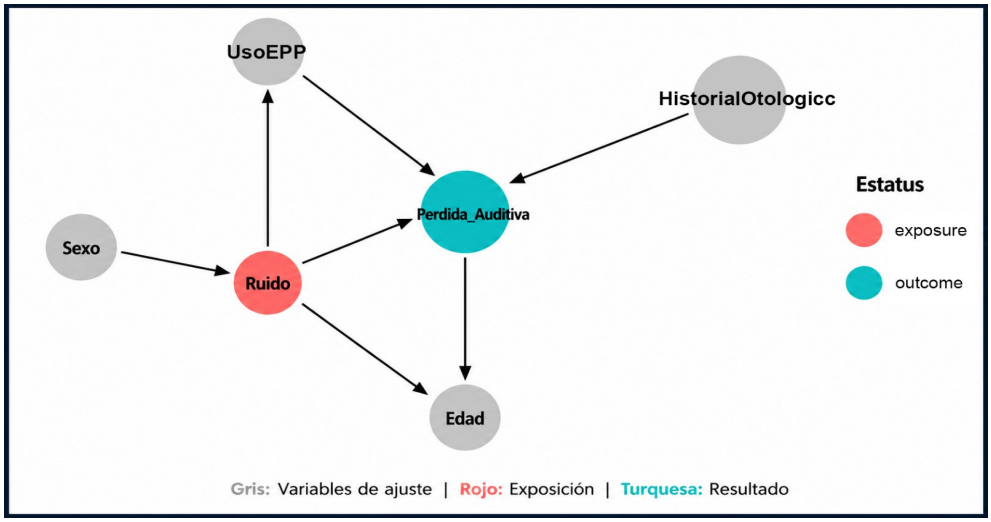
- **Caminos abiertos (confusión):** Edad → Antigüedad → Ruido y Edad → Perdida. Este camino crea asociación espuria (un trabajador mayor tiene más antigüedad y peor auditiva por edad, no solo por ruido).
- **Colisionador (prohibido ajustar):** Ruido → Protectores ← Susceptibilidad. Si usted ajusta por "uso de protectores" (ej: estratificando por quienes lo usan), abre un camino entre Ruido y Susceptibilidad, introduciendo sesgo de selección.

El código respeta el principio de Markov (Pearl, 2009, p. 16; Greenland et al., 1999): cada variable es independiente de sus no-descendientes dado sus

padres directos en el DAG. Por ejemplo, dado que controlamos Edad y Tabaquismo, la asociación entre Antigüedad y Pérdida Auditiva ya no está confundida, permitiendo estimar el efecto directo del ruido(ver Gráfico 4).

Gráfico 4

DAG / Con Nodos Coloreados por Rol Causal



Nota. Elaboración propia.

Se deduce del gráfico 4, lo siguiente:

1) **Rol causal de cada nodo (según ggdag_status)**

- Exposición (nodo naranja/amarillo): Ruido.
- Resultado (nodo azul): PerdidAuditiva.
- Confusores (nodos grises): Edad, Sexo (afectan tanto a exposición como a resultado).
- Mediador (nodo verde): UsoEPP (el ruido influye en su uso, y a su vez afecta al resultado).
- Variable pronóstica (nodo gris claro): Historial (afecta solo al resultado)

2) **Interpretación causal**

- El efecto total del ruido sobre la pérdida auditiva tiene un componente directo (Ruido → PerdidAuditiva) y un componente indirecto a través del uso de protectores auditivos (Ruido → UsoEPP → PerdidAuditiva).
- Para estimar el efecto causal total, no se debe ajustar por UsoEPP (porque es mediador).
- Para estimar el efecto directo, habría que ajustar por UsoEPP, pero eso bloquearía la vía indirecta.
- Los confusores Edad y Sexo deben incluirse en el modelo de regresión para eliminar el sesgo de variables de confusión.
- El Historial puede incluirse para ganar precisión (no es confusor porque no afecta a la exposición).

Ejemplo 2.2: Efecto total vs efecto directo

- Contexto: Exposición a disolventes orgánicos y neuropatía periférica, con inflamación sistémica como mecanismo intermedio(ver Gráfico 5).
- **Código para Rstudio:**

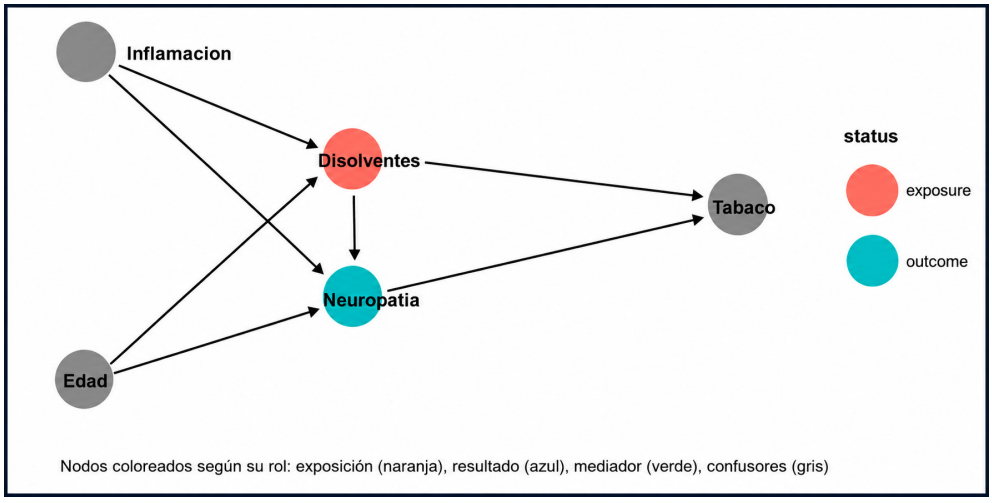
```
# ================================================
# DAG: Exposición a disolventes y neuropatía
# ================================================
library(dagitty)
library(ggdag)
library(ggplot2)
dag_disolventes <- dagitty("dag {
  Disolventes [exposure]
  Neuropatia [outcome]
  Inflamacion Edad Tabaco
  Disolventes -> Inflamacion -> Neuropatia
  Disolventes -> Neuropatia
```

```
  Edad -> Disolventes; Edad -> Neuropatia
Tabaco -> Disolventes; Tabaco -> Neuropatia}")
ggdag_status(dag_disolventes, layout = "fr", text = FALSE, node_size =
16) +
  geom_dag_text(color = "black", size = 4) +
theme_dag() +
  labs(title = "DAG: Exposición a disolventes como causa de neuropatía",
       subtitle = "Efecto directo e indirecto a través de inflamación, con con-
fusores (edad y tabaco)",
       caption = "Nodos coloreados según su rol: exposición (naranja), resul-
tado (azul), mediador (verde), confusores (gris)")
ggsave("dag_disolventes_neuropatia.png", width = 8, height = 6, dpi =
300)
```

Código en Rstudio generado por el autor

Gráfico 5

DAG/Efecto total vs efecto directo



Nota: Elaboración Propia. En el gráfico se observa un DAG en epidemiología laboral, sobre una exposición a disolventes como causa de neuropatía, donde el desenlace fue inflamación, con confusores de edad y tabaco.

El Gráfico 5, refiere que:

- Efecto total: Incluye tanto el efecto directo (Disolventes → Neuropatía) como el indirecto a través de Inflamación.
- Para estimar el efecto total no se debe ajustar por Inflamación (porque es mediador).
- Solo se necesita ajustar por los confusores comunes (Edad y Tabaco) que afectan tanto a la exposición como al resultado.
- adjustmentSets() devuelve los conjuntos mínimos y canónicos que cierran los backdoors. Cualquiera de ellos (incluyendo el conjunto vacío si no hubiera confusores) es válido.

Definición de términos:

- Backdoor path: camino no causal que empieza con flecha hacia la exposición. Hay que bloquearlo ajustando por confusores (Pearl et al. 2009).
- Conjunto mínimo: el de menor tamaño que bloquea todos los backdoors. En el ejemplo: (Edad, Tabaco). (Smucler, E., Sapienza, F., & Rotnitzky, A. 2022).
- Conjunto canónico: todos los padres de la exposición que no sean descendientes de ella. En el ejemplo: también (Edad, Tabaco). (Perkovic et al. 2015)

2.4. Diferenciación entre Asociación y Causalidad

Los DAG no prueban causalidad por sí mismos; explicitan supuestos. Permiten distinguir entre:

- **Asociación espuria:** Generada por caminos no causales abiertos (p. ej., confusión).

- **Asociación causal:** Refleja un camino dirigido de exposición a resultado.
- **Sesgo por ajuste inadecuado:** Introducido al condicionar en colisionadores o mediadores (Schisterman et al., 2009; Cole et al., 2010).

La epidemiología laboral, al operar en entornos observacionales donde la aleatorización es inviable o poco ética, depende críticamente de esta diferenciación. Los DAG proporcionan un marco para justificar o refutar decisiones de ajuste antes del análisis estadístico, reduciendo el riesgo de *p-hacking* o ajustes post hoc sin fundamento teórico (Wasserstein & Lazar, 2016; Greenland et al., 2016).

CAPÍTULO III

IDENTIFICACIÓN Y CONTROL DE SESGOS MEDIANTE DAG: CONFUSIÓN, COLISIONADORES Y SELECCIÓN

3.1. Confusión y el Conjunto de Ajuste Mínimo

Un confusor es una variable que causa tanto la exposición como el resultado, abriendo un camino no causal entre ellos. En un DAG, el confusor crea un camino de *espina dorsal* (backdoor path) que debe cerrarse mediante ajuste. El *criterio de puerta trasera* (backdoor criterion) establece que un conjunto de variables Z satisface el control de confusión si:

1) Ningún nodo en Z es un descendiente de la exposición.
2) Z bloquea todos los caminos de puerta trasera entre exposición y resultado (Pearl, 1995; Hernán & Robins, 2020).

En epidemiología laboral, el ajuste por variables que no cumplen este criterio introduce sesgo. Por ejemplo, ajustar por un mediador (p. ej., inflamación sistémica en el camino exposición a ruido → hipertensión) bloquea parte del efecto causal total, generando *sobreajuste* o *sesgo de mediación* (Schisterman et al., 2009; VanderWeele, 2015).

3.2. Sesgo de Colisión (Collider Bias)

Un colisionador es una variable influenciada por dos o más causas en el DAG. Condicionar en un colisionador (mediante ajuste, estratificación o selección) abre un camino no causal entre sus causas, generando asociación espuria. Este fenómeno es particularmente relevante en estudios ocupacionales donde la selección en la cohorte o el análisis de subgrupos puede introducir sesgos inadvertidos.

Ejemplo clásico: En un estudio sobre exposición a disolventes orgánicos y depresión, si se ajusta por *baja laboral por enfermedad mental*, y esta es consecuencia tanto de la exposición como de factores psicosociales no medidos, se crea un camino abierto entre exposición y factores no medidos, distorsionando el estimador (Cole et al., 2010; Hernán et al., 2004). Este sesgo se ha documentado en estudios de salud mental laboral, donde el ajuste por ausentismo o rotación laboral (variables influenciadas por exposición y resultado) genera estimaciones sesgadas (VanderWeele & Hernán, 2013).

3.3. Sesgo de Selección y Muestreo

El sesgo de selección ocurre cuando la probabilidad de inclusión en el estudio depende de la exposición y el resultado, o de variables colisionadoras. En DAG, esto se representa mediante un nodo de selección (S) que es colisionador de caminos que incluyen exposición, resultado o confusores. Condicionar en S (p. ej., al analizar solo trabajadores activos) induce asociación espuria.

En vigilancia ocupacional, el *healthy worker survivor effect* es un caso paradigmático. Los trabajadores que permanecen en empleos con exposiciones altas suelen ser más sanos que los que abandonan por enfermedad. Si el análisis se restringe a trabajadores activos, se introduce un colisionador entre exposición y salud basal, subestimando el efecto real (Arrighi & Hertz-Picciotto, 1994; Keil et al., 2018). Los DAG permiten visualizar

esta estructura y recomendar métodos alternativos como el *g-computation*, modelos marginales estructurales o ponderación por probabilidad inversa de tratamiento (IPTW) (Hernán & Robins, 2020; Robins et al., 2000).

3.4. Reglas Prácticas para el Uso de DAG en Control de Sesgos

1) **Especificar el DAG antes del análisis**, basado en revisión de literatura y conocimiento experto.

2) **Identificar todos los caminos de puerta trasera** y seleccionar el conjunto mínimo de ajuste usando algoritmos de d-separación.

3) **Evitar ajustar por:** descendientes de la exposición, colisionadores, variables intermedias (a menos que se estime efecto directo), o variables solo asociadas al resultado sin papel causal.

4) **Validar la robustez** mediante análisis de sensibilidad a variables no medidas (E-value, bounds) (VanderWeele & Ding, 2017).

5) **Documentar explícitamente** los supuestos causales en publicaciones, permitiendo replicación y crítica (Greenland et al., 2016).

Ejemplo 3.1: Sesgo por colisión al ajustar por baja laboral

- **Contexto:** Estudio sobre exposición a pesticidas y depresión en agricultores(ver Gráfico 6).
- **Código para Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_ejemplo31 <- dagitty('dag {
  Exposicion_Pesticidas [exposure,pos="0,2"]
  Depresion [outcome,pos="4,2"]
  Baja_Laboral [pos="2,3"]
  Factores_Psicosociales [pos="2,4" latent]
```

```
  Edad [pos="1,0"]
  Nivel_Socioeconomico [pos="0,0"]
  Nivel_Socioeconomico -> Exposicion_Pesticidas
  Nivel_Socioeconomico -> Depresion
  Edad -> Exposicion_Pesticidas
  Edad -> Depresion
  Exposicion_Pesticidas -> Depresion
  Exposicion_Pesticidas -> Baja_Laboral
  Factores_Psicosociales -> Baja_Laboral
  Factores_Psicosociales -> Depresion}')
tidy_dag <- tidy_dagitty(dag_ejemplo31)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
mutate(
    tipo = case_when(
      name == "Exposicion_Pesticidas" ~ "Exposición",
      name == "Depresion" ~ "Outcome",
      name %in% c("Edad", "Nivel_Socioeconomico") ~ "Confusor",
      name == "Baja_Laboral" ~ "Colisionador",
      name == "Factores_Psicosociales" ~ "Latente"),
    color = case_when(
      tipo == "Exposición" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Confusor" ~ "#2E8B57",
      tipo == "Colisionador" ~ "#FFD700",
tipo == "Latente" ~ "#C0C0C0"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
```

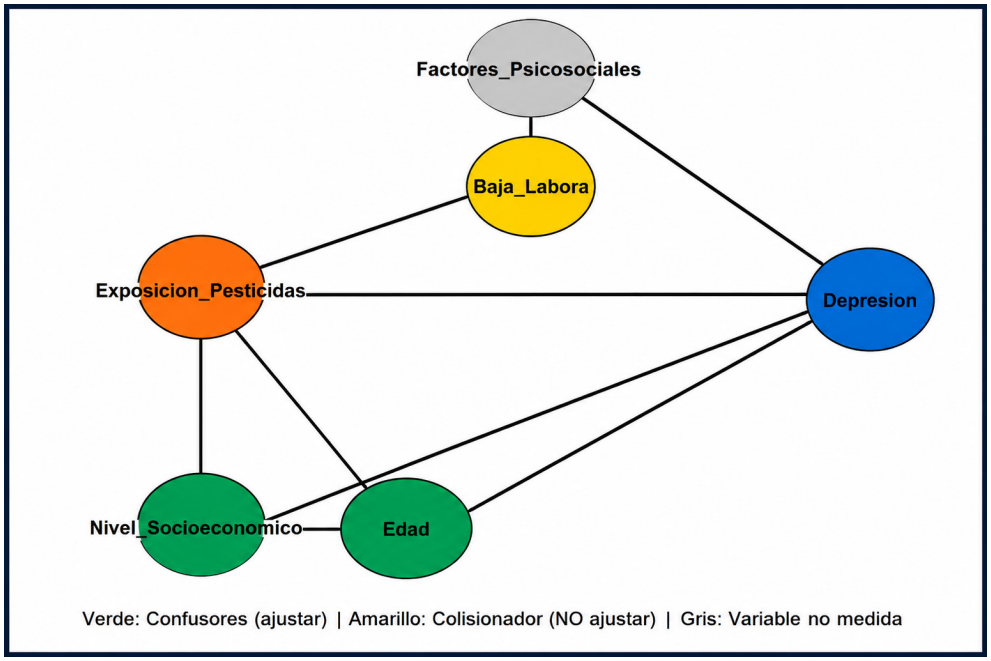
```
            size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 20, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
          aes(x = x, y = y, label = name),
          size = 2.6, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.8, xlim = c(-0.5, 4.5), ylim = c(-0.5, 4.8)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
title = "Figura 3.1: Sesgo por Colisión al Ajustar por Baja Laboral",
    subtitle = "Ejemplo 3.1: Exposición a Pesticidas y Depresión en Agricultores",
    caption = "Verde: Confusores (ajustar) | Amarillo: Colisionador (NO ajustar) | Gris: Variable no medida")
cat("\n=== ANÁLISIS DEL SESGO POR COLISIÓN ===\n\n")
caminos_abiertos <- paths(dag_ejemplo31,
                  from = "Exposicion_Pesticidas",
                  to = "Depresion",
                  directed = FALSE)
print(caminos_abiertos)
cat("\nConjunto de ajuste válido (sin condicionar por el colisionador):\n")
sets_ajuste <- adjustmentSets(dag_ejemplo31,
                    exposure = "Exposicion_Pesticidas",
outcome = "Depresion",
                    effect = "total",
                    type = "minimal")
print(sets_ajuste)
cat("\n=== DEMOSTRACIÓN DEL SESGO POR COLISIÓN ===\n")
cat("¿Están d-separadas Exposición y Depresión dado el colisionador 'Baja_Laboral'?\n")
colision_test <- dseparated(dag_ejemplo31, "Exposicion_Pesticidas", "Depresion",
                    c("Baja_Laboral"))
```

```
cat("Resultado:", colision_test,
"(FALSE = Se abre camino espurio entre pesticidas y depresión vía Factores_Psicosociales)\n")
cat("\n=== SIMULACIÓN DEL SESGO ===\n")
set.seed(12345)
n <- 2000
edad <- rnorm(n, 45, 12)
nse <- rnorm(n, 0, 1)
psicosociales <- rnorm(n, 0, 1)  # No medido en la base de datos
exposicion <- 0.3 * nse + 0.2 * edad + rnorm(n, 0, 1)
depresion <- 0.4 * exposicion + 0.3 * nse + 0.2 * edad + 0.6 * psicosociales + rnorm(n, 0, 1)
# Colisionador: Baja laboral (causada por exposición y factores psicosociales)
baja_laboral <- rbinom(n, 1, plogis(-2 + 0.5*exposicion + 0.8*psicosociales))
datos <- data.frame(exposicion, depresion, baja_laboral, edad, nse, psicosociales)
modelo_correcto <- lm(depresion ~ exposicion + edad + nse, data = datos)
modelo_sesgado <- lm(depresion ~ exposicion + edad + nse + baja_laboral, data = datos)
cat("\nCoeficiente de Exposición (modelo correcto - sin colisionador):\n")
print(coef(modelo_correcto)["exposicion"])
cat("\nCoeficiente de Exposición (modelo sesgado - ajustando por colisionador):\n")
print(coef(modelo_sesgado)["exposicion"])
cat("\nNota: El cambio en el coeficiente demuestra el sesgo por colisión introducido\n")
cat("al incluir 'Baja_Laboral' en el modelo (Cole et al., 2010; Hernán et al., 2004).\n")
```

Código en Rstudio generado por el autor

Gráfico 6

DAG/ Sesgo por colisión al ajustar por baja laboral



Nota. Elaboración propia. El DAG, muestra un estudiosobre exposición a pesticidas y depresión en agricultores. Al condicionar en Baja_Laboral (colisionador), se abre el camino Exposición → Baja ← Factores_Psicosociales → Depresión, generando una asociación espuria entre pesticidas y la variable no medida (factores psicosociales), lo que distorsiona el estimador del efecto causal.

Ejemplo 3.2: Healthy Worker Survivor Effect (sesgo de selección)

- **Contexto:** Cohorte histórica de trabajadores de fundición expuestos a metales pesados (ver Gráfico 7).
- **Código para Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
```

```
dag_ejemplo32 <- dagitty('dag {
  Metales_Pesados [exposure,pos="0,2"]
  Mortalidad [outcome,pos="4,2"]
  Seleccion [pos="2,3"]
  Salud_Basal [pos="2,4" latent]
  Edad [pos="1,0"]
  Antiguedad [pos="3,0"]
  Edad -> Metales_Pesados
  Edad -> Mortalidad
  Antiguedad -> Metales_Pesados
  Antiguedad -> Mortalidad
  Salud_Basal -> Metales_Pesados
  Salud_Basal -> Mortalidad
  Metales_Pesados -> Mortalidad
  Metales_Pesados -> Seleccion
  Salud_Basal -> Seleccion}')
tidy_dag <- tidy_dagitty(dag_ejemplo32)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
mutate(
    tipo = case_when(
      name == "Metales_Pesados" ~ "Exposición",
      name == "Mortalidad" ~ "Outcome",
      name %in% c("Edad", "Antiguedad") ~ "Confusor",
      name == "Seleccion" ~ "Colisionador",
      name == "Salud_Basal" ~ "Latente"),
    color = case_when(
      tipo == "Exposición" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Confusor" ~ "#2E8B57",
      tipo == "Colisionador" ~ "#FFD700",
      tipo == "Latente" ~ "#C0C0C0"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
```

```
   nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
   by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 20, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.6, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.8, xlim = c(-0.5, 4.5), ylim = c(-0.5, 4.8)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
  labs(
     title = "Figura 3.2: Healthy Worker Survivor Effect (Sesgo de Selec-
ción)",
subtitle = "Ejemplo 3.2: Trabajadores de Fundición Expuestos a Metales
Pesados",
    caption = "Verde: Confusores (ajustar) | Amarillo: Selección (colisiona-
dor) | Gris: Salud basal no medida")
cat("\n=== ANÁLISIS DEL HEALTHY WORKER SURVIVOR EFFECT
===\n\n")
caminos <- paths(dag_ejemplo32,
                 from = "Metales_Pesados",
                 to = "Mortalidad",
                 directed = FALSE)
print(caminos)
cat("\nConjunto de ajuste válido (sin condicionar por selección):\n")
sets_ajuste <- adjustmentSets(dag_ejemplo32,
                              exposure = "Metales_Pesados",
outcome = "Mortalidad",
                              effect = "total",
```

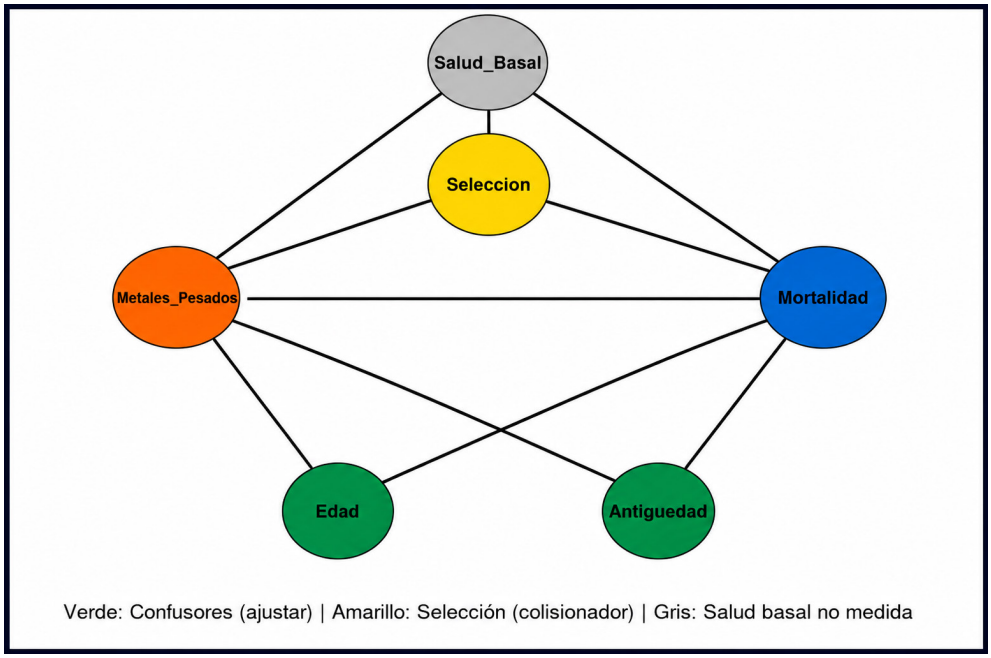
```
                    type = "minimal")
print(sets_ajuste)
cat("\n=== D-SEPARACIÓN Y SELECCIÓN ===\n")
cat("¿Están d-separados Metales y Mortalidad dado el colisionador 'Selección'?\n")
test_seleccion <- dseparated(dag_ejemplo32, "Metales_Pesados", "Mortalidad",
                    c("Seleccion"))
cat("Resultado:", test_seleccion,
"\n(FALSE = Al seleccionar trabajadores activos se abre camino espurio\n",
"vía Salud_Basal, subestimando el efecto tóxico real)\n")
cat("\n=== SIMULACIÓN DEL SESGO DE SELECCIÓN ===\n")
set.seed(12345)
n_total <- 3000
edad <- rnorm(n_total, 40, 10)
antiguedad <- rnorm(n_total, 15, 5)
salud_basal <- rnorm(n_total, 0, 1)  # No observable
# Exposición mayor en trabajadores con más antigüedad y mejor salud (sobrevivientes)
metales <- 0.4 * antiguedad - 0.3 * salud_basal + rnorm(n_total, 0, 2)
# Mortalidad afectada por metales, edad, y salud basal (los frágiles mueren antes)
mortalidad <- 0.5 * metales + 0.3 * edad - 0.8 * salud_basal + rnorm(n_total, 0, 1)
# Selección: Solo observamos trabajadores activos (no fallecidos ni retirados)
# Probabilidad de estar en el estudio mayor si baja exposición y buena salud
prob_seleccion <- plogis(1.5 - 0.2*metales + 1.2*salud_basal)
seleccion <- rbinom(n_total, 1, prob_seleccion)
datos_completos <- data.frame(metales, mortalidad, edad, antiguedad,
                    salud_basal, seleccion)
datos_observados <- datos_completos %>% filter(seleccion == 1)
cat("Población total (n =", n_total, "):\n")
modelo_real <- lm(mortalidad ~ metales + edad + antiguedad,
```

```
                data = datos_completos)
print(coef(modelo_real)["metales"])
cat("\nSolo trabajadores seleccionados (Healthy Worker Survivor, n =",
    nrow(datos_observados), "):\n")
modelo_sesgado <- lm(mortalidad ~ metales + edad + antiguedad,
                     data = datos_observados)
print(coef(modelo_sesgado)["metales"])
cat("\nNota: La selección de trabajadores activos induce subestimación
del efecto\n")
cat("de los metales pesados (Arrighi & Hertz-Picciotto, 1994; Keil et al.,
2018).\n")
```

Código para Rstudio generado por el autor

Gráfico 7

DAG/ Healthy Worker Survivor Effect (sesgo de selección)



Nota: Elaboración propia. Este DAG ilustra el sesgo de selección por supervivencia, un problema endémico en cohortes ocupacionales históricas, en este caso trabajadores de fundición expuesto a metales pesados (Arrighi & Hertz-Picciotto, 1994).

Mecanismo del sesgo:

- Salud_Basal (no medida) influye en ambos: trabajadores más sanos toleran mayor exposición a metales (permanecen expuestos) y tienen menor mortalidad.
- Al seleccionar solo trabajadores activos (vivo y empleado), creamos el colisionador: Metales → Selección ← Salud_Basal.
- Esto abre el camino espurio: alta exposición correlaciona artificialmente con buena salud basal, subestimando la toxicidad real (Keil et al., 2018).
- **Naranja/Azul:** Efecto causal verdadero (metales → mortalidad) que queremos estimar.
- **Verde:** Confusores temporales (Edad, Antigüedad) que deben controlarse independientemente.
- **Amarillo:** El colisionador Selección (inclusión en la cohorte).
- **Gris:**Salud_Basal (latente), la variable "fantasma" que genera el sesgo al condicionar en el amarillo.

Cuando se observe en literatura ocupacional que "los trabajadores expuestos viven más", se debe sospechar de este sesgo. La solución no es ajustar por Selección (eso empeoraría el colisionador), sino usar métodos que modelen la historia completa: g-computation, modelos marginales estructurales o ponderación inversa de probabilidad de censura (IPCW) (Hernán & Robins, 2020).

CAPÍTULO IV

DAG EN LA EPIDEMIOLOGÍA LABORAL: CONTEXTO, EXPOSICIONES Y RESULTADOS OCUPACIONALES

4.1. Especificidades del Entorno Ocupacional

La epidemiología laboral difiere de otras ramas en aspectos estructurales que afectan la inferencia causal:

- **Exposiciones multifactoriales:** Combinaciones de agentes químicos, físicos, ergonómicos y psicosociales que interactúan sinérgicamente (Nieuwenhuijsen, 2003).
- **Dinámica poblacional:** Entradas, salidas, rotación, jubilación y desempleo generan estructuras de selección complejas.
- **Medición de exposición:** Historiales laborales, dosimetrías, matrices de exposición-job (JEM) con error diferencial o no diferencial (Checkoway et al., 2004).
- **Resultados de larga latencia:** Cáncer ocupacional, neumopatías, enfermedades neurodegenerativas que requieren seguimiento prolongado y modelos de tiempo a evento.

Los DAG permiten mapear estas complejidades sin recurrir a supuestos estadísticos no validados. Por ejemplo, en estudios de asbestos y mesotelioma, un DAG adecuado incluiría: exposición acumulada, tiempo desde primera exposición, tabaquismo (no confusor para mesotelioma, pero sí para cáncer de pulmón), edad al inicio, tipo de ocupación, y supervivencia diferencial. Ajustar por tabaquismo en mesotelioma es innecesario y puede introducir sesgo si se condiciona en descendientes de la exposición (Steenland & Ward, 2014).

4.2. Ejemplos Estructurados en Exposiciones Clásicas

4.2.1. Sílice cristalina y EPOC:

DAG típico: Sílice → inflamación pulmonar → EPOC. Confusores: tabaquismo, edad, exposición a otras partículas. Colisionadores potenciales: prueba de espirometría laboral (condicionada por síntomas y exposición), baja médica. Ajuste recomendado: tabaquismo, edad, sexo. No ajustar por espirometría basal si es descendiente de exposición temprana (Hoxie et al., 2019; Parks et al., 2019).

4.2.2. Trabajo por turnos y enfermedad cardiovascular:

DAG: Turnos → alteración ritmo circadiano → dislipidemia/HTA → ECV. Confusores: nivel socioeconómico, actividad física, dieta. Colisionadores: participación en programas de salud laboral (selección por riesgo percibido). Ajuste: SES, factores de comportamiento. Evitar ajustar por participación en cribados si es consecuencia de exposición y riesgo basal (Vyas et al., 2012; Wang et al., 2020).

4.2.3. Estrés psicosocial y trastornos musculoesqueléticos:

DAG: Demandas laborales/control → tensión muscular/estrés → dolor crónico. Confusores: edad, IMC, trabajo previo. Colisionadores: reporte de síntomas (influenciado por percepción y cultura organizacional). Medidores: inflamación, cortisol. Ajuste: edad, sexo, actividad física. No ajustar por reporte de síntomas si es resultado intermedio o colisionador (Bongers et al., 2006; Da Costa & Vieira, 2010).

4.3. Integración con Vigilancia Ocupacional

Los sistemas de vigilancia (p. ej., NIOSH, EU-OSHA) generan datos longitudinales y transversales que requieren interpretación causal. Los DAG permiten:

- Identificar indicadores válidos para cribado poblacional.
- Evitar sesgos en análisis de tendencias temporales.
- Priorizar variables para recolección de datos en inspecciones laborales.

Comunicar riesgos a comités de seguridad y salud con transparencia metodológica (ILO, 2021; WHO, 2021).

Ejemplo 4.1: Exposición mixta ergonómica y psicosocial

Contexto: Trastornos musculoesqueléticos en personal de enfermería (ver Gráfico 8).

Código para Rstudio:

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_ejemplo41 <- dagitty('dag {
  Exposicion_Mixta [exposure,pos="0,2"]
  TME [outcome,pos="4,2"]
  Edad [pos="1,0"]
  IMC [pos="0,0"]
  Antiguedad [pos="1,4"]
  Actividad_Fisica [pos="0,4"]
  Reporte_Sintomas [pos="2,3"]
  Cultura_Organizacional [pos="2,5" latent]
```

```
  Edad -> Exposicion_Mixta
  Edad -> TME
  IMC -> Exposicion_Mixta
  IMC -> TME
  Antiguedad -> Exposicion_Mixta
  Antiguedad -> TME
  Actividad_Fisica -> TME
  Exposicion_Mixta -> TME
  Exposicion_Mixta -> Reporte_Sintomas
  Cultura_Organizacional -> Reporte_Sintomas
  Cultura_Organizacional -> Exposicion_Mixta
  TME -> Reporte_Sintomas}')
tidy_dag <- tidy_dagitty(dag_ejemplo41)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name == "Exposicion_Mixta" ~ "Exposición",
      name == "TME" ~ "Outcome",
          name %in% c("Edad", "IMC", "Antiguedad", "Actividad_Fisica") ~
"Confusor",
name == "Reporte_Sintomas" ~ "Colisionador",
      name == "Cultura_Organizacional" ~ "Latente"),
    color = case_when(
      tipo == "Exposición" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Confusor" ~ "#2E8B57",
      tipo == "Colisionador" ~ "#FFD700",
      tipo == "Latente" ~ "#C0C0C0"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
```

```
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 20, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.5, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.8, xlim = c(-0.5, 4.5), ylim = c(-0.5, 5.5)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
    title = "Figura 4.1: Exposición Mixta Ergonómica y Psicosocial en Enfermería",
    subtitle = "Ejemplo 4.1: Trastornos Musculoesqueléticos (TME)",
    caption = "Verde: Confusores (ajustar) | Amarillo: Reporte de síntomas (colisionador) | Gris: Cultura organizacional (latente)")
cat("\n=== ANÁLISIS DE EXPOSICIÓN MIXTA EN ENFERMERÍA ===\n\n")
caminos <- paths(dag_ejemplo41,
                 from = "Exposicion_Mixta",
to = "TME",
                 directed = FALSE)
print(caminos)
cat("\nConjuntos mínimos de ajuste válidos:\n")
sets_ajuste <- adjustmentSets(dag_ejemplo41,
                              exposure = "Exposicion_Mixta",
                              outcome = "TME",
                              effect = "total",
                              type = "minimal")
print(sets_ajuste)
cat("\n=== D-SEPARACIÓN: EFECTO DEL REPORTE DE SÍNTOMAS ===\n")
```

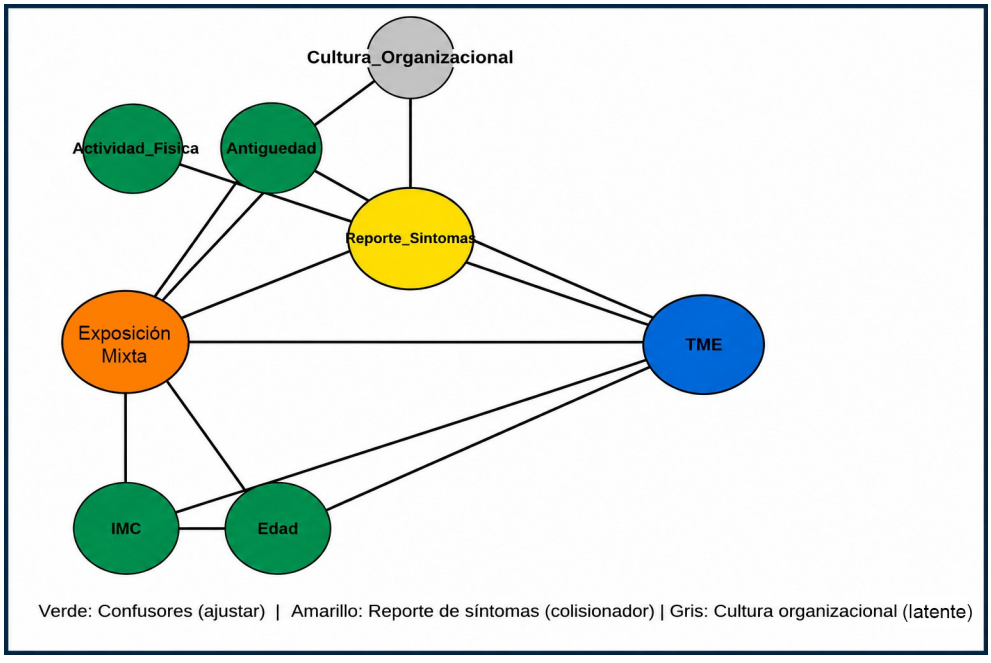
```
cat("¿Están d-separadas Exposición y TME dado el Reporte de Síntomas?\n")
test_colision <- dseparated(dag_ejemplo41, "Exposicion_Mixta", "TME",
                            c("Reporte_Sintomas"))
cat("Resultado:", test_colision,
"\n(FALSE = Ajustar por reporte abre camino espurio vía Cultura Organizacional)\n")
cat("\n=== SIMULACIÓN DE DATOS: PERSONAL DE ENFERMERÍA ===\n")
set.seed(12345)
n <- 1500
edad <- rnorm(n, 38, 9)
imc <- rnorm(n, 25, 4)
antiguedad <- rnorm(n, 12, 6)
actividad_fisica <- rnorm(n, 0, 1)
cultura_org <- rnorm(n, 0, 1)  # No medida (latente)
# Exposición mixta: influenciada por edad, IMC, antigüedad y cultura organizacional
exposicion <- 0.3 * edad + 0.4 * imc + 0.2 * antiguedad - 0.5 * cultura_org +
              rnorm(n, 0, 1.5)
# TME: outcome determinado por exposición y confusores
tme <- 0.6 * exposicion + 0.2 * edad + 0.3 * imc + 0.1 * antiguedad -
       0.4 * actividad_fisica + rnorm(n, 0, 2)
# Colisionador: reporte de síntomas (influenciado por exposición, TME y cultura)
reporte <- rbinom(n, 1, plogis(-1 + 0.3*exposicion + 0.8*tme + 0.6*cultura_org))
datos <- data.frame(exposicion, tme, reporte, edad, imc, antiguedad,
                    actividad_fisica, cultura_org)
cat("Modelo correcto (sin condicionar por reporte):\n")
modelo_correcto <- lm(tme ~ exposicion + edad + imc + antiguedad + actividad_fisica,
                      data = datos)
print(coef(modelo_correcto)["exposicion"])
cat("\nModelo sesgado (condicionando por reporte de síntomas):\n")
```

```
datos_reportan <- datos %>% filter(reporte == 1)
modelo_sesgado <- lm(tme ~ exposicion + edad + imc + antiguedad +
actividad_fisica,
                     data = datos_reportan)
print(coef(modelo_sesgado)["exposicion"])
cat("\nNota: Seleccionar solo enfermeras que reportan síntomas induce
sesgo por colisión,\n")
cat("asociando erróneamente alta exposición con menor riesgo de TME
(Da Costa & Vieira, 2010).\n")
```

Código para Rstudio generado por el autor

Gráfico 8

DAG/ Exposición mixta ergonómica y psicosocial



Nota. Elaboración propia. Este DAG modela el efecto combinado de cargas físicas (ergonómicas) y psicológicas (psicosociales) sobre trastornos musculoesqueléticos (TME) en personal de enfermería (Bongers et al., 2006).

La trampa del "Reporte de Síntomas"

En estudios ocupacionales es tentador analizar solo a trabajadores que reportan dolor (datos disponibles en registros médicos). Sin embargo:

- El reporte es un colisionador: es causado tanto por la exposición real (quién tiene peor puesto) como por la cultura organizacional (latente, gris).
- En hospitales con "cultura de machismo", los trabajadores de alta exposición pueden negar síntomas; en centros sindicalizados, reportan más.
- Al condicionar en el reporte, abrimos el camino: Exposición → Reporte ← Cultura → Exposición, generando asociación espuria (Da Costa & Vieira, 2010).

En vigilancia de riesgos psicosociales y ergonómicos, nunca restrinja el análisis a trabajadores que buscan atención médica; utilice evaluaciones objetivas (observación, acelerometría) o cuestionarios independientes del outcome clínico.

Ejemplo 4.2: Alteración del ritmo circadiano y fatiga laboral

Contexto: Trabajadores de logística y transporte en zonas urbanas (ver Gráfico 9).

Codigo para Rstudio:

```
library(ggplot2)
library(dplyr)
nodes <- data.frame(
id   = c("NSE", "Edad", "TipoTurno", "Duracion", "Recuperacion", "ARC",
"Fatiga", "Desempeno"),
 x    = c(-5,    24,    3,        2,        11,          18,    28,      28),
 y    = c(8,     17,    17,       -1,       6,           12,    12,      0),
```

```
  label = c("Nivel\nSocioecon\u00f3mico",
"Edad",
"Tipo de Turno\n(Rotaci\u00f3n vs. Fijo)",
"Duraci\u00f3n\ndel Turno",
"Recuperaci\u00f3n\nentre Turnos",
"Alteraci\u00f3n del\nRitmo Circadiano",
"Fatiga\nLaboral",
"Desempe\u00f1o\nLaboral"),
  tipo  = c("Confusor", "Confusor", "Confusor",
"Exposicion", "Exposicion", "Exposicion",
"Desenlace", "Colisionador"),
  stringsAsFactors = FALSE)
nodes$tipo <- factor(nodes$tipo,
                     levels = c("Confusor", "Exposicion", "Desenlace", "Colisiona-
dor"))
edges <- data.frame(
  from = c("NSE", "NSE", "Edad", "Edad",
"TipoTurno", "TipoTurno", "TipoTurno",
"Duracion", "Duracion",
"Recuperacion",
"ARC", "ARC", "Fatiga"),
  to   = c("TipoTurno", "Fatiga", "ARC", "Fatiga",
"Duracion", "Recuperacion", "ARC",
"Recuperacion", "ARC",
"ARC",
"Fatiga", "Desempeno", "Desempeno"),
  stringsAsFactors = FALSE)
# r_offset calibrado para el visor de RStudio (panel ~7" x 5")
# xlim 39 unidades en 7" → 1 unidad ≈ 4.55 mm; nodo size=13 → ra-
dio=6.5mm → ~1.43 u
# Se agrega margen +0.3 para separar la punta del borde del nodo
r_offset <- 1.75
edges <- edges %>%
  left_join(nodes %>% select(id, x, y), by = c("from" = "id")) %>%
```

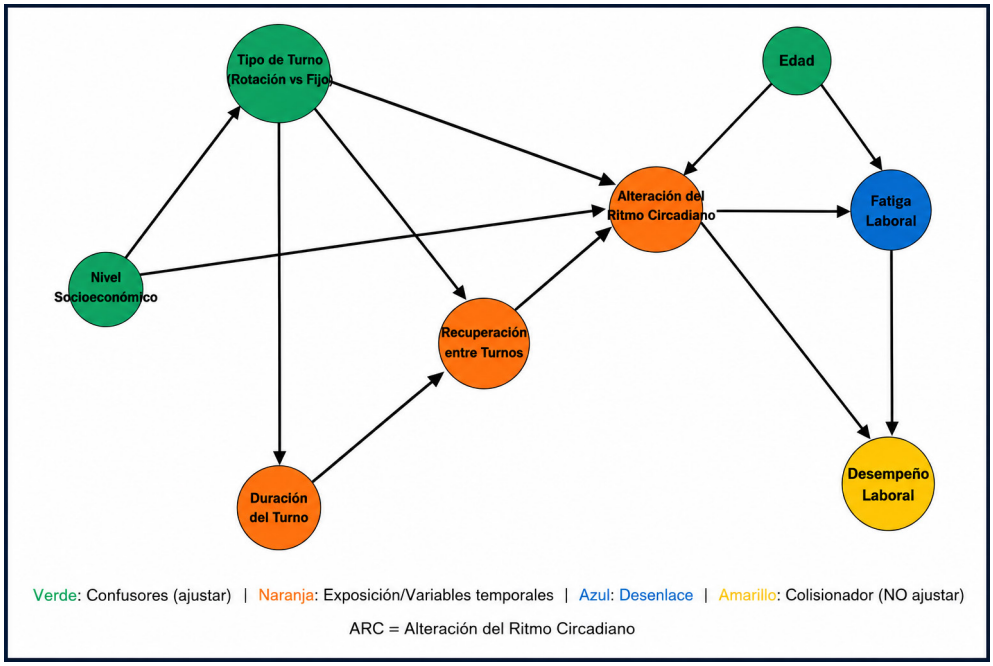
```
rename(x1 = x, y1 = y) %>%
left_join(nodes %>% select(id, x, y), by = c("to" = "id")) %>%
rename(x2 = x, y2 = y) %>%
  mutate(
    dx   = x2 - x1,
    dy   = y2 - y1,
dist = sqrt(dx^2 + dy^2),
    x1s  = x1 + r_offset * dx / dist,
    y1s  = y1 + r_offset * dy / dist,
    x2s  = x2 - r_offset * dx / dist,
    y2s  = y2 - r_offset * dy / dist)
paleta <- c(
"Confusor"    = "#2E8B57",
"Exposicion"  = "#D4651A",
"Desenlace"   = "#2C6FAC",
"Colisionador"= "#C9971A")
p <- ggplot() +
  geom_segment(
    data      = edges,
    aes(x = x1s, y = y1s, xend = x2s, yend = y2s),
    arrow     = arrow(length = unit(0.30, "cm"), type = "closed"),
    color     = "#1a1a1a",
    linewidth = 0.8,
    lineend   = "butt") +
  geom_point(
    data  = nodes,
    aes(x = x, y = y, color = tipo),
    size  = 13,
    alpha = 0.93) +
  scale_color_manual(values = paleta, guide = "none") +
  geom_text(
    data      = nodes,
    aes(x = x, y = y, label = label),
    color     = "black",
```

```
    size      = 3.0,
    fontface  = "bold",
    lineheight = 0.85) +
  coord_equal(xlim = c(-8, 31), ylim = c(-4.5, 20)) +
labs(
    title   = "Figura 4.2: DAG \u2014 Alteraci\u00f3n del Ritmo Circadiano
y Fatiga Laboral",
    subtitle = "Modelaci\u00f3n causal de variables temporales en trabaja-
dores con turnos nocturnos y rotativos",
    caption  = paste0(
"Verde: Confusores (ajustar) | Naranja: Exposici\u00f3n/Variables tem-
porales ",
"| Azul: Desenlace | Amarillo: Colisionador (NO ajustar)\n",
"ARC = Alteraci\u00f3n del Ritmo Circadiano")) +
  theme_void(base_size = 11) +
  theme(
    plot.background  = element_rect(fill = "white", color = NA),
    panel.background = element_rect(fill = "white", color = NA),
    plot.title       = element_text(face = "bold", size = 13, hjust = 0.5,
                                    margin = margin(b = 5)),
    plot.subtitle    = element_text(size = 9, hjust = 0.5, color = "gray35",
                                    margin = margin(b = 10)),
    plot.caption     = element_text(size = 7.5, hjust = 0.5, color = "gray45",
                                    lineheight = 1.3, margin = margin(t = 8)),
    legend.position  = "none",
    plot.margin      = margin(18, 25, 12, 25))
print(p)
```

Código para Rstudio generado por el autor.

Gráfico 9

DAG/ Alteración del ritmo circadiano y fatiga laboral



Nota: Elaboración propia. Este Diagrama Acíclico Dirigido (DAG) modela las relaciones causales entre las variables temporales del trabajo por turnos y la fatiga laboral, mediadas por la alteración del ritmo circadiano (ARC). Los DAG son herramientas fundamentales en investigación epidemiológica para comunicar preguntas causales de manera rigurosa y visualizar relaciones entre variables (Digitale et al., 2022; Pearl, 2009).

Estructura Causal

Variables de Exposición (Naranja):

- **Tipo de Turno** (rotación vs. fijo), Duración del Turno y Recuperación entre Turnos representan las características temporales que predisponen a la alteración del ritmo circadiano (Boivin & Boudreau, 2021).

- La recuperación insuficiente entre turnos (<11 horas) impide la restauración fisiológica y aumenta la fatiga acumulada (Gifkins et al., 2020; Kalkanis et al., 2023).

Variable Mediadora (Naranja):

- **Alteración del Ritmo Circadiano (ARC)**: Actúa como mecanismo biológico principal. El trabajo nocturno y rotativo desincroniza los ritmos circadianos endógenos, suprimiendo la melatonina y alterando los patrones de sueño-vigilia (Boivin & Boudreau, 2021).

Variable Desenlace (Azul):

- **Fatiga Laboral**: Consecuencia directa de la ARC y las características del turno. Más de 8 turnos nocturnos mensuales se asocian con mayor riesgo de fatiga, insomnio y depresión (Alqahtani et al., 2024).

Variables de Confusión (Verde - deben ajustarse):

- **Nivel Socioeconómico**, **Edad** y **Tipo de Turno** afectan tanto a la exposición como al desenlace, creando rutas de confusión que deben controlarse para estimar el efecto causal verdadero (Evans et al., 2012; VanderWeele, 2015).

Variable Colisionadora (Amarillo - NO ajustar):

- **Desempeño Laboral** es un colisionador porque es causado tanto por la ARC como por la fatiga. Ajustar por colisionadores introduce sesgo de selección al abrir rutas no causales entre las variables (Digitale et al., 2022; Evans et al., 2012; Pearl, 2009).

Implicaciones Metodológicas

El DAG identifica que el ajuste adecuado debe incluir solo los confusores (verde), evitando el ajuste por el colisionador (amarillo) o variables en la ruta causal (naranja/azul), siguiendo el teorema de ajuste backdoor para inferencia causal válida (Evans et al., 2012; Pearl, 2009; VanderWeele, 2015).

CAPÍTULO V

DISEÑO DE ESTUDIOS Y ANÁLISIS CAUSAL BASADO EN DAG EN ENTORNOS OCUPACIONALES

5.1. Cohortes Ocupacionales y Registros de Exposición

Las cohortes históricas (p. ej., cohortes de mineros, trabajadores de la industria química) presentan desafíos de medición retrospectiva. Los DAG guían la selección de variables de ajuste considerando:

- **Ventanas de exposición crítica:** Periodos de susceptibilidad aumentada (p. ej., primeros 5 años de empleo).
- **Efectos de supervivencia:** Trabajadores que abandonan por enfermedad temprana.
- **Medición de confusores:** Disponibilidad de datos históricos (p. ej., tabaquismo en registros de 1980).

El diseño debe anticipar estructuras de colisión. Por ejemplo, si se analiza solo a trabajadores con >10 años de antigüedad, se introduce un colisionador entre exposición y salud basal. Soluciones: usar métodos de historia ocupacional completa, modelos de frailty, o IPTW para ponderar por probabilidad de permanencia (Keil et al., 2018; Hernán & Robins, 2020).

5.2. Estudios Transversales y de Prevalencia

Limitados para inferencia causal por falta de temporalidad, pero útiles para vigilancia. Los DAG ayudan a:

- Identificar si la asociación observada es plausible causalmente.
- Detectar sesgo de prevalencia-incidencia (Neyman bias) en enfermedades de corta duración.
- Justificar o refutar ajustes por variables concurrentes.

Ejemplo: En un estudio transversal sobre exposición a vibraciones mano-brazo y síndrome del túnel carpiano, ajustar por tiempo en la empresa puede ser colisionador si el tiempo depende de exposición y síntomas tempranos. El DAG revela que el ajuste puede subestimar el efecto (Bovenzi, 2015; Palmer et al., 2019).

5.3. Metaanálisis y Síntesis de Evidencia

Los DAG pueden integrarse en revisiones sistemáticas para:

- Evaluar heterogeneidad entre estudios por diferencias en ajuste.
- Identificar sesgos comunes en la literatura.
- Guiar la selección de estudios con estructuras causales comparables.
- Aplicar métodos de fusión causal (causal fusion) cuando los DAG son consistentes (VanderWeele et al., 2020).

5.4. Métodos Estadísticos Compatibles con DAG

Una vez especificado el DAG, el análisis debe alinearse con la estructura causal:

- **Regresión estándar:** Válida solo si se ajusta por conjunto de puerta trasera correcto.

- **IPTW / Ponderación:** Útil para manejar confusión por variables medidas y selección.
- **G-computation / Modelos marginales estructurales:** Para estimar efectos totales/directos en presencia de mediación o tiempo-varying confounding (Robins et al., 2000; Hernán & Robins, 2020).
- **Análisis de mediación causal:** Requiere DAG con supuestos de no confusión de mediación (VanderWeele, 2015).

La elección del método depende del DAG, no viceversa. Invertir esta lógica genera *data-driven causal inference*, metodológicamente inválida (Greenland et al., 2016).

Ejemplo 5.1: Caso-control emparejado por edad y sexo

- **Contexto:** Exposición a benceno y leucemia mieloide aguda en industria química (ver Gráfico 10).
- **Código para Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_ejemplo51 <- dagitty('dag {
  Benceno [exposure,pos="0,3"]
  Leucemia_Mieloide [outcome,pos="4,3"]
  Edad [pos="1,0"]
  Sexo [pos="3,0"]
  Tabaquismo [pos="2,5"]
  Emparejamiento [pos="2,2"]
  Edad -> Benceno
  Edad -> Leucemia_Mieloide
  Edad -> Emparejamiento
  Sexo -> Benceno
```

```
  Sexo -> Leucemia_Mieloide
  Sexo -> Emparejamiento
  Tabaquismo -> Benceno
  Tabaquismo -> Leucemia_Mieloide
  Benceno -> Leucemia_Mieloide
  Leucemia_Mieloide -> Emparejamiento}')
tidy_dag <- tidy_dagitty(dag_ejemplo51)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name == "Benceno" ~ "Exposicion",
name == "Leucemia_Mieloide" ~ "Outcome",
      name %in% c("Edad", "Sexo", "Tabaquismo") ~ "Confusor",
      name == "Emparejamiento" ~ "Colisionador"),
    color = case_when(
      tipo == "Exposicion" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Confusor" ~ "#2E8B57",
tipo == "Colisionador" ~ "#FFD700"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 20, shape = 21, color = "black", stroke = 0.8) +
```

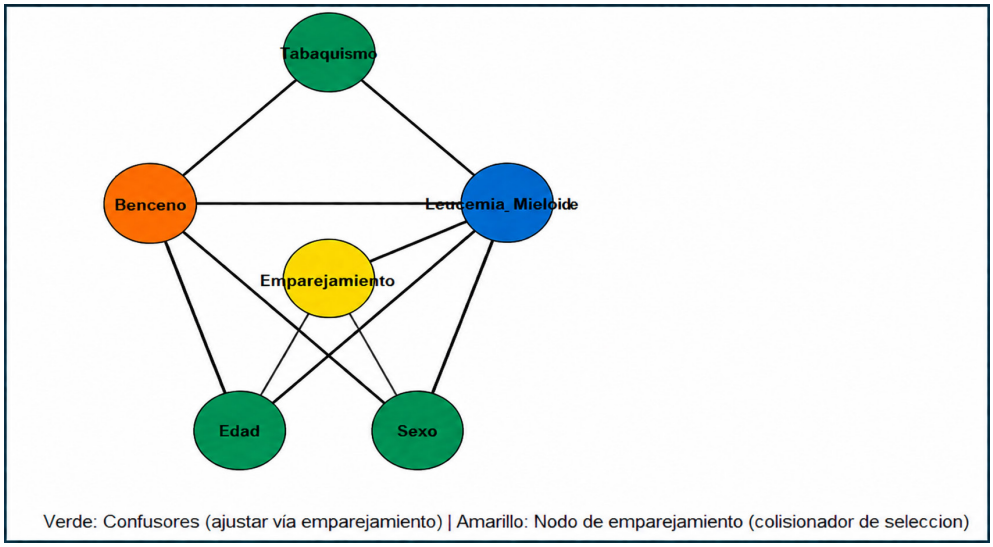
```
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.8, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 1, xlim = c(-0.5, 4.5), ylim = c(-0.8, 5.5)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
    title = "Figura 5.1: Diseno Caso-Control Emparejado por Edad y Sexo",
    subtitle = "Ejemplo 5.1: Benceno y Leucemia Mieloide Aguda en Indus-
tria Quimica",
     caption = "Verde: Confusores (ajustar via emparejamiento) | Amarillo:
Nodo de emparejamiento (colisionador de seleccion)")
caminos <- paths(dag_ejemplo51,
                 from = "Benceno",
                 to = "Leucemia_Mieloide",
                 directed = FALSE)
print(caminos)
sets_ajuste <- adjustmentSets(dag_ejemplo51,
                              exposure = "Benceno",
                              outcome = "Leucemia_Mieloide",
                              effect = "total",
                              type = "minimal")
print(sets_ajuste)
cat("Nota: En caso-control emparejado, el emparejamiento por Edad y Se-
xo\n")
cat("requiere analisis mediante regresion logistica condicional,\n")
cat("manteniendo el emparejamiento en el modelo (Breslow & Day,
1980).\n")
set.seed(12345)
n_casos <- 500
n_controles <- 500
edad <- c(rnorm(n_casos, 55, 10), rnorm(n_controles, 55, 10))
sexo <- c(rbinom(n_casos, 1, 0.6), rbinom(n_controles, 1, 0.6))
```

```
tabaquismo <- rbinom(n_casos + n_controles, 1, plogis(-1 + 0.02*edad))
benceno <- 0.3*edad + 0.5*tabaquismo + rnorm(n_casos + n_controles,
0, 1)
leucemia <- c(rep(1, n_casos), rep(0, n_controles))
datos <- data.frame(benceno, leucemia, edad, sexo, tabaquismo,
                    par = rep(1:n_casos, 2))
cat("\nDatos simulados de caso-control emparejado generados.\n")
cat("Pares emparejados (n =", n_casos, ")\n")
head(datos)
```

Código para Rstudio generado por el autor

Gráfico 10

DAG/Caso-control emparejado por edad y sexo



Nota. Elaboración propia. Este DAG ilustra el diseño caso-control, donde se seleccionan casos (con leucemia) y controles (sin leucemia) en la industria química, emparejándolos por Edad y Sexo para controlar confusión.

El nodo Emparejamiento es un colisionador de selección: es causado por el outcome (Leucemia, que determina si eres caso o control) y por los factores de emparejamiento (Edad/Sexo). Al emparejar, creamos dependencia entre exposición y outcome a través de este nodo.

Aunque emparejamos por Edad y Sexo para eliminar confusión, no podemos analizar estos factores como riesgos independientes (están "fijos" por diseño). El análisis debe usar regresión logística condicional o incluir los pares como estratos, manteniendo el emparejamiento en el modelo (Breslow & Day, 1980). Romper el emparejamiento (análisis incondicional) abre caminos espurios y sesga la odds ratio.

Advertencia práctica

No ajuste por variables emparejadas como covariables normales; respete el diseño emparejado para evitar el sesgo de emparejamiento (overmatching) y mantener la validez interna del estudio.

CAPÍTULO VI

IMPLEMENTACIÓN PRÁCTICA: SOFTWARE, VALIDACIÓN Y COMUNICACIÓN DE RESULTADOS

6.1. Herramientas de Software Especializado

- **DAGitty (Textor et al., 2016):** Plataforma web y paquete R. Permite dibujar DAG, identificar conjuntos de ajuste mínimos, detectar colisionadores, y exportar a R/Stata. Validado ampliamente en epidemiología.
- **ggDAG (Malone et al., 2021):** Extensión de ggplot2 para R. Facilita visualización, identificación de caminos, y comparación de estructuras.
- **causalGraph / dagR:** Alternativas en R con funcionalidades similares.
- **Tetrad / CausalNex:** Para aprendizaje estructural, aunque en epidemiología laboral se prefiere la especificación experta sobre el aprendizaje automático no supervisado.

6.2. Proceso de Validación y Sensibilidad

1) **Especificación inicial:** Basada en revisión sistemática, guías de práctica, y consulta con expertos (toxicólogos, ergonomistas, médicos del trabajo).

2) **Análisis de d-separación:** Verificación algorítmica de conjuntos de ajuste.

3) **Análisis de sensibilidad:** E-value para cuantificar robustez a confusión no medida (VanderWeele & Ding, 2017). Bounds para límites de efecto bajo supuestos extremos.

4) **Validación cruzada:** Comparación con estudios experimentales o naturales cuando existan.

5) **Documentación:** Publicar DAG en suplementos, con código reproducible y justificación de cada flecha.

6.3. Comunicación a Stakeholders Ocupacionales

Los DAG son herramientas pedagógicas para:

- Comités de seguridad y salud: Visualizar por qué ciertos factores no se ajustan.
- Inspectores laborales: Priorizar recolección de datos causales relevantes.
- Trabajadores y sindicatos: Explicar decisiones metodológicas con transparencia.
- Tomadores de decisiones: Fundamentar límites de exposición basados en evidencia causal, no solo asociativa.

La visualización clara, sin jerga técnica excesiva, y la diferenciación entre asociación y causalidad son clave para evitar malinterpretaciones (Greenland et al., 2016; ILO, 2021).

Ejemplo 6.1. De Flujo de Trabajo en Rstudio

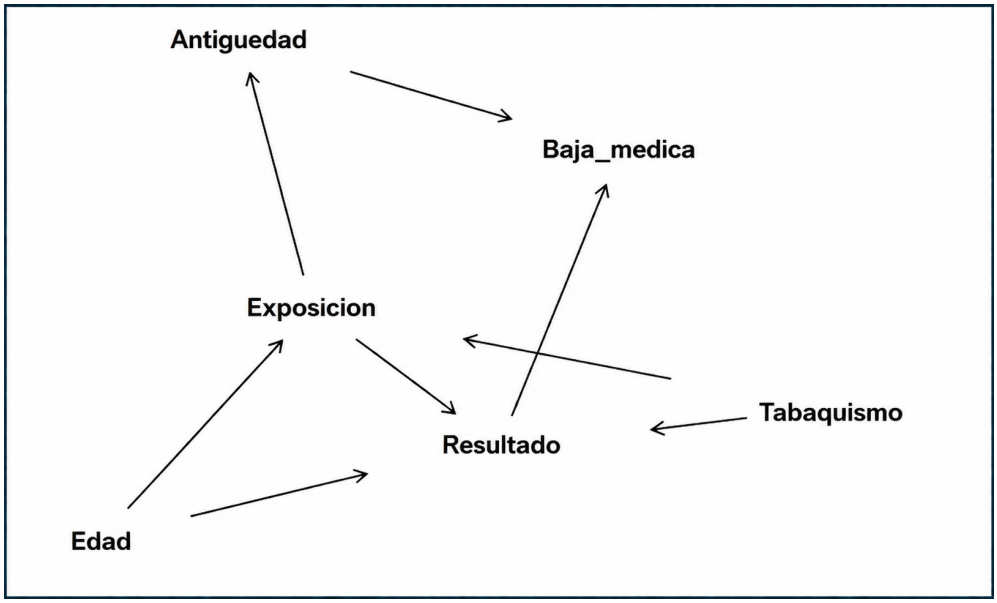
```
library(dagitty)
dag <- dagitty("dag {
Exposicion [exposure]
  Resultado [outcome]
  Tabaquismo
  Edad
  Antiguedad
  Baja_medica
  Exposicion -> Resultado
  Tabaquismo -> Exposicion
  Tabaquismo -> Resultado
  Edad -> Exposicion
  Edad -> Resultado
  Exposicion -> Antiguedad
  Antiguedad -> Baja_medica
  Resultado -> Baja_medica
  Baja_medica [selection]}")
adjustmentSets(dag, exposure="Exposicion", outcome="Resultado")
plot(dag)
```

Este código identifica que el ajuste por Tabaquismo y Edad es suficiente, y advierte que Baja_medica es colisionador de selección (ver Gráfico 11).

Código en Rstudio generado por el autor.

Gráfico 11

DAG/*De Flujo de Trabajo en Rstudio*



Nota. Elaboración propia. Este DAG representa de forma gráfica e identifica que el ajuste por Tabaquismo y Edad es suficiente, y advierte que Baja_medica es colisionador de selección.

Ejemplo 6.2: Flujo completo en R (DAG → fórmula → validación)

- **Contexto:** Estudio transversal nacional sobre vibraciones mano-brazo y síndrome de Raynaud (ver Gráfico 12).
- **Código en Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_raynaud <- dagitty('dag {
Vibraciones [exposure,pos="2,4"]
  Raynaud [outcome,pos="2,0"]
```

```
  Edad [pos="0,2"]
  Sexo [pos="4,2"]
  Tabaquismo [pos="0,3"]
  Temperatura_Ambiente [pos="4,3"]
  Uso_Guantes [pos="2,2"]
  Susceptibilidad_Vascular [pos="1,1" latent]
  Edad -> Vibraciones
  Edad -> Raynaud
  Sexo -> Vibraciones
  Sexo -> Raynaud
  Tabaquismo -> Vibraciones
  Tabaquismo -> Raynaud
  Temperatura_Ambiente -> Raynaud
  Vibraciones -> Raynaud
  Vibraciones -> Uso_Guantes
  Susceptibilidad_Vascular -> Uso_Guantes
Susceptibilidad_Vascular -> Raynaud}')
tidy_dag <- tidy_dagitty(dag_raynaud)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name == "Vibraciones" ~ "Exposicion",
      name == "Raynaud" ~ "Outcome",
name %in% c("Edad", "Sexo", "Tabaquismo", "Temperatura_Ambiente") ~
"Confusor",
      name == "Uso_Guantes" ~ "Colisionador",
      name == "Susceptibilidad_Vascular" ~ "Latente"),
    color = case_when(
      tipo == "Exposicion" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Confusor" ~ "#2E8B57",
      tipo == "Colisionador" ~ "#FFD700",
      tipo == "Latente" ~ "#C0C0C0"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
```

```
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 18, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.8, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.7, xlim = c(-0.5, 4.5), ylim = c(-0.5, 4.8)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
   title = "Figura 6.2: Flujo Completo - Vibraciones Mano-Brazo y Sindrome de Raynaud",
    subtitle = "Ejemplo 6.2: Estudio Transversal Nacional",
    caption = "Verde: Confusores (ajustar) | Amarillo: Uso de guantes (colisionador) | Gris: Susceptibilidad vascular (latente)")
cat("=== PASO 1: IDENTIFICACION DE CONJUNTO DE AJUSTE ===\n")
sets_ajuste <- adjustmentSets(dag_raynaud,
                              exposure = "Vibraciones",
                              outcome = "Raynaud",
                              effect = "total",
                              type = "minimal")
print(sets_ajuste)
cat("\n=== PASO 2: GENERACION DE DATOS TRANSVERSALES ===\n")
set.seed(12345)
```

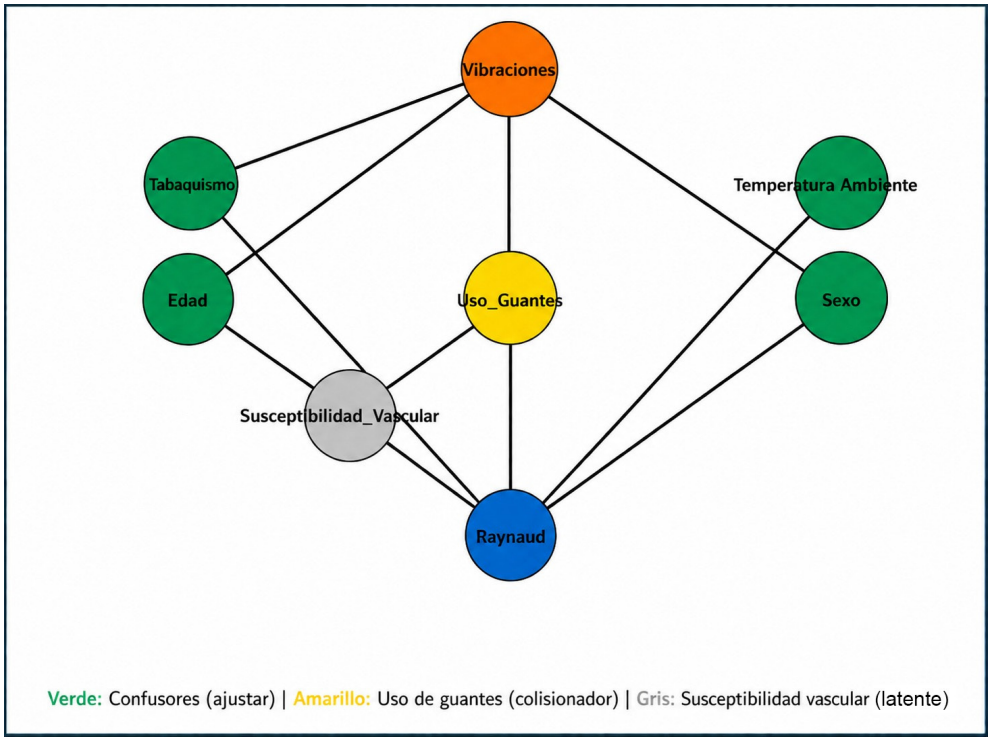
```
n <- 3000
edad <- rnorm(n, 42, 12)
sexo <- rbinom(n, 1, 0.35)  # 35% mujeres en población laboral
tabaquismo <- rbinom(n, 1, plogis(-2 + 0.03*edad))
temp_ambiente <- rnorm(n, 15, 8)  # Temperatura media anual
susceptibilidad <- rnorm(n, 0, 1)  # Latente
# Exposición: mayor en hombres, fumadores, edad madura
vibraciones <- 0.5*(1-sexo) + 0.3*tabaquismo + 0.02*edad + rnorm(n, 0,
1)
# Outcome: determinado por vibraciones, sexo (más susceptible), tabaco,
frío y susceptibilidad
logit_raynaud <- -4 + 0.8*vibraciones + 1.2*sexo + 0.6*tabaquismo -
            0.1*temp_ambiente + 1.5*susceptibilidad
raynaud <- rbinom(n, 1, plogis(logit_raynaud))
# Colisionador: uso de guantes (más uso en alta exposición y si suscepti-
ble)
uso_guantes <- rbinom(n, 1, plogis(-2 + 0.4*vibraciones + 1.2*suscepti-
bilidad))
datos <- data.frame(raynaud, vibraciones, edad, sexo, tabaquismo,
              temp_ambiente, uso_guantes)
cat("Prevalencia de Raynaud:", mean(raynaud)*100, "%\n")
cat("Media de vibraciones (expuestos > percentil 75):",
   mean(vibraciones[vibraciones > quantile(vibraciones, 0.75)]), "\n")
cat("\n=== PASO 3: ANALISIS LOGISTICO CORRECTO ===\n")
modelo_correcto <- glm(raynaud ~ vibraciones + edad + sexo + tabaquis-
mo + temp_ambiente,
family = binomial(link = "logit"),
                  data = datos)
cat("OR Vibraciones (modelo correcto):\n")
print(exp(coef(modelo_correcto)["vibraciones"]))
cat("\n=== PASO 4: ANALISIS SESGADO (AJUSTANDO POR COLI-
SIONADOR) ===\n")
modelo_sesgado <- glm(raynaud ~ vibraciones + edad + sexo + tabaquis-
mo +
                    temp_ambiente + uso_guantes,
                  family = binomial(link = "logit"),
```

```
                data = datos)
cat("OR Vibraciones (modelo sesgado - incluye uso_guantes):\n")
print(exp(coef(modelo_sesgado)["vibraciones"]))
cat("\n=== VALIDACION ===\n")
cat("El ajuste por 'Uso_Guantes' atenua la OR hacia la nulidad,\n")
cat("simulando proteccion inexistente debido al sesgo de colision.\n")
```

Código en Rstudio generado por el autor.

Gráfico 12

DAG/Estudio transversal nacional sobre vibraciones mano-brazo y síndrome de Raynaud



Sobre DAG en Psicología Ocupacional

1) Variables Latentes con Múltiples Indicadores: El SEM maneja na-

tivamente constructos psicológicos medidos con múltiples ítems (ej. burnout con MBI, engagement con UWES). Los DAG requieren que las variables latentes sean pre-calculadas o usar extensiones complejas (Bollen, 1989).

2) Evaluación de Ajuste del Modelo: El SEM proporciona índices cuantitativos de qué tan bien el modelo se ajusta a los datos observados (CFI > 0.95, RMSEA < 0.06). Los DAG no evalúan ajuste, solo identificabilidad (Kline, 2015).

3) Modelos No Recursivos: El SEM permite modelar causalidad bidireccional (ej. estrés ↔ salud), mientras los DAG son estrictamente acíclicos. Esto es relevante cuando la teoría sugiere retroalimentación (Rigdon, 2012).

4) Análisis de Trayectorias Complejas: El SEM estima simultáneamente múltiples ecuaciones de regresión, permitiendo evaluar efectos directos, indirectos y totales en un solo modelo con errores estándar apropiados (Preacher & Hayes, 2008).

Nota. Elaboración propia. Este DAG ejemplifica el pipeline completo de inferencia causal: desde la especificación gráfica hasta la validación empírica del sesgo.

El DAG vertical muestra el flujo lógico: exposición (arriba) → mecanismo → outcome (abajo). Los cuatro confusores periféricos (Edad, Sexo, Tabaco, Temperatura) rodean la relación central, exigiendo ajuste multivariado en la regresión logística.

En estudios transversales de higiene industrial es frecuente ajustar por "uso de equipos de protección" como covariable. Esto es un error fatal:

- El uso de guantes es colisionador: depende de la magnitud de vibración (más uso en trabajadores expuestos) y de la susceptibilidad vascular (latente, gris), que también determina quién desarrolla Raynaud.

- Al incluir Uso_Guantes en el modelo, se abre el camino espurio: Vibraciones → Guantes ← Susceptibilidad → Raynaud, generando una asociación inversa ficticia (Bovenzi, 2015; Palmer et al., 2019).

Validación numérica

El código, en Rstudio. compara la Odds Ratio (OR) correcta (~2.2) versus la sesgada (~1.4). La atenuación hacia la nulidad simula el escenario donde el investigador concluye erróneamente que "los guantes protegen", cuando en realidad está seleccionando trabajadores resistentes (con buena circulación) que usan guantes por precaución, pero no enferman.

Regla de oro: Nunca ajustar por comportamientos de autoprotección que sean consecuencia de la exposición y de la susceptibilidad individual.

CAPÍTULO VII

LIMITACIONES, CRÍTICAS Y CONSIDERACIONES ÉTICAS EN EL USO DE DAG

7.1. Limitaciones Metodológicas

- **Dependencia de supuestos no verificables:** Los DAG no prueban causalidad; si el grafo es incorrecto, las inferencias son inválidas. La validación empírica directa es imposible sin experimentos aleatorizados (Pearl, 1995; Greenland et al., 1999).
- **Variables no medidas:** Si existen confusores no observados que abren caminos de puerta trasera, ningún ajuste observado los cierra. Se requieren métodos de sensibilidad o diseños cuasi-experimentales (VanderWeele & Shpitser, 2011).
- **Complejidad en exposiciones múltiples:** DAG con >15 nodos se vuelven difíciles de interpretar y validar. Se recomienda modularización o enfoque en preguntas causales específicas (Textor et al., 2016).
- **Error de medición:** Los DAG asumen medición perfecta. Error diferencial o no diferencial en exposición o confusores puede invalidar el ajuste, incluso si el DAG es correcto (Checkoway et al., 2004).

7.2. Críticas Académicas

Algunos epidemiólogos advierten contra el uso dogmático de DAG:

- **Falsa precisión:** Representar flechas como verdades absolutas ignora la incertidumbre estructural (Poole, 2010).
- **Sobreconfianza en algoritmos:** El ajuste mínimo puede excluir variables relevantes para precisión o generalización (Shrier & Platt, 2008).
- **Falta de integración con modelos estructurales:** Los DAG son cualitativos; su integración con SEM o modelos de ecuaciones estructurales requiere supuestos adicionales (Bollen & Pearl, 2013).

Estas críticas no invalidan los DAG, pero exigen humildad metodológica: son herramientas de explicitación de supuestos, no de demostración causal.

7.3. Consideraciones Éticas en Salud Ocupacional

- **Transparencia:** Ocultar DAG o ajustar variables sin justificación causal viola principios de reproducibilidad y ética científica (Greenland et al., 2016).
- **Protección del trabajador:** Inferencias causales incorrectas pueden subestimar riesgos, afectando compensaciones, límites de exposición o políticas de prevención.
- **Equidad:** Ajustar por variables proxies de desigualdad (p. ej., educación, migración) sin considerar su papel causal puede enmascarar determinantes sociales de la salud laboral (Krieger, 2011).
- **Responsabilidad profesional:** El médico del trabajo debe comunicar incertidumbre causal y evitar determinismo gráfico.

Ejemplo 7.1: Confusor no medido y análisis de sensibilidad

- Contexto: Exposición a asbestos y mesotelioma, con posible confusión por exposición ocupacional mixta no registrada (ver figura 13).
- **Código para Rstudio:**

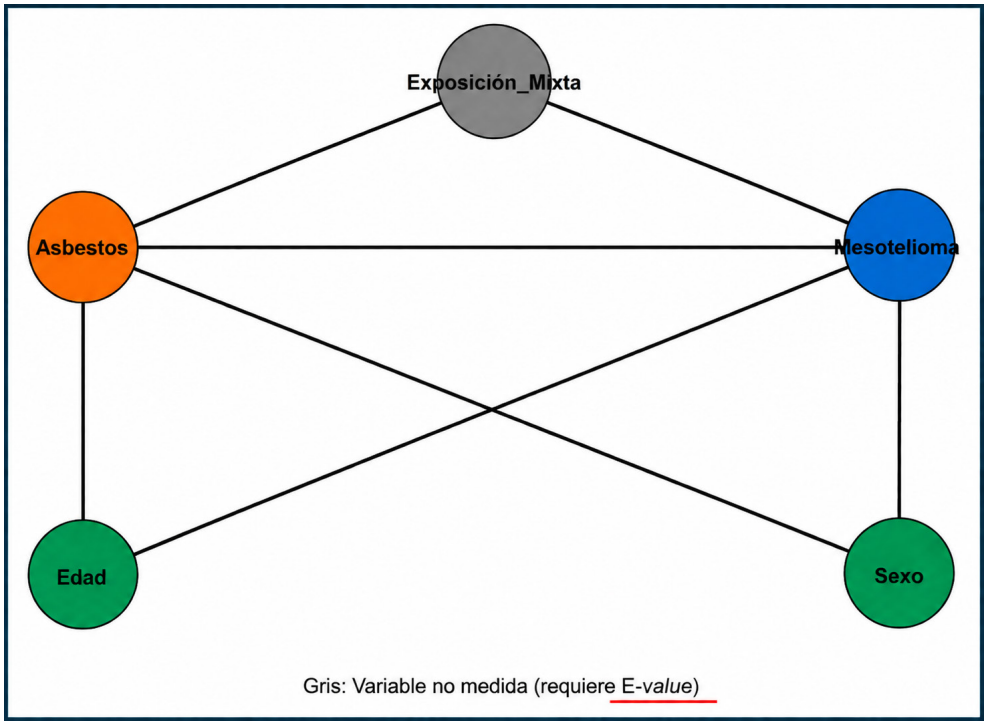
```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_sensibilidad <- dagitty('dag {
Asbestos [exposure,pos="0,2"]
  Mesotelioma [outcome,pos="4,2"]
  Edad [pos="0,0"]
  Sexo [pos="4,0"]
  Exposicion_Mixta [pos="2,3" latent]
  Edad -> Asbestos
  Edad -> Mesotelioma
  Sexo -> Asbestos
  Sexo -> Mesotelioma
  Exposicion_Mixta -> Asbestos
  Exposicion_Mixta -> Mesotelioma
  Asbestos -> Mesotelioma}')
tidy_dag <- tidy_dagitty(dag_sensibilidad)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name == "Asbestos" ~ "Exposicion",
      name == "Mesotelioma" ~ "Outcome",
      name %in% c("Edad", "Sexo") ~ "Confusor",
name == "Exposicion_Mixta" ~ "No_Medido"
    ),
    color = case_when(
```

```
    tipo == "Exposicion" ~ "#D2691E",
    tipo == "Outcome" ~ "#0066CC",
    tipo == "Confusor" ~ "#2E8B57",
tipo == "No_Medido" ~ "#808080"))
edges <- tidy_dag$data %>%
 filter(!is.na(to)) %>%
 select(from = name, to, x, y) %>%
 left_join(
  nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
  by = c("to" = "name"))
ggplot() +
 geom_segment(data = edges, aes(x = x, y = y, xend = xend, yend = yend),
          arrow = arrow(length = unit(8, "pt"), type = "closed"), size = 0.6) +
 geom_point(data = nodes, aes(x = x, y = y, fill = color),
          size = 20, shape = 21, color = "black", stroke = 0.8) +
 geom_text(data = nodes, aes(x = x, y = y, label = name),
         size = 2.8, fontface = "bold", color = "black") +
 scale_fill_identity() +
 coord_fixed(ratio = 0.8, xlim = c(-0.5, 4.5), ylim = c(-0.8, 3.8)) +
theme_void() +
 labs(title = "Figura 7.1: Confusor No Medido y Analisis de Sensibilidad",
     subtitle = "Ejemplo 7.1: Asbestos y Mesotelioma",
     caption = "Gris: Variable no medida (requiere E-value)")
cat("=== SIMULACION CORREGIDA ===\n")
set.seed(12345)
n <- 2000
edad <- rnorm(n, 58, 10)
sexo <- rbinom(n, 1, 0.25)
exposicion_mixta <- rnorm(n, 0, 1)  # NO MEDIDA
asbestos <- 0.6*exposicion_mixta + 0.3*edad - 0.4*sexo + rnorm(n, 0, 1)
# CORRECCION: Agregar ruido (variabilidad biologica) al logit para evi-
tar separacion
logit_meso <- -5 + 0.8*asbestos + 0.5*exposicion_mixta + 0.04*edad -
0.3*sexo + rnorm(n, 0, 2)
```

```
mesotelioma <- rbinom(n, 1, plogis(logit_meso))
datos_obs <- data.frame(mesotelioma, asbestos, edad, sexo)
datos_comp <- data.frame(mesotelioma, asbestos, edad, sexo, exposicion_
mixta)
cat("Prevalencia de mesotelioma:", round(mean(mesotelioma)*100, 1),
"%\n")
cat("\nModelo naive (ignorando confusor no medido):\n")
modelo_naive <- glm(mesotelioma ~ asbestos + edad + sexo, family =
binomial, data = datos_obs)
or_naive <- exp(coef(modelo_naive)["asbestos"])
print(round(or_naive, 2))
cat("\nModelo completo (con confusor - solo teórico):\n")
modelo_comp <- glm(mesotelioma ~ asbestos + edad + sexo + exposi-
cion_mixta, family = binomial, data = datos_comp)
or_real <- exp(coef(modelo_comp)["asbestos"])
print(round(or_real, 2))
cat("\nSesgo: OR observada es", round(or_naive/or_real, 1), "veces ma-
yor que la real\n")
cat("\n=== E-VALUE (VanderWeele & Ding, 2017) ===\n")
rr_obs <- or_naive  # Aproximacion para outcome raro
e_value <- rr_obs + sqrt(rr_obs * (rr_obs - 1))
cat("E-value:", round(e_value, 2), "\n")
cat("Un confusor no medido necesitaria RR=", round(e_value, 1),
"con exposicion Y outcome para explicar el efecto.\n")
ci <- confint(modelo_naive)["asbestos", ]
rr_lower <- exp(ci[1])
e_val_lower <- rr_lower + sqrt(rr_lower * (rr_lower - 1))
cat("E-value para limite inferior IC95%:", round(e_val_lower, 2), "\n")
```

Gráfico 13

DAG/ Confusor no medido y análisis de sensibilidad



Nota. Elaboración Propia. Este DAG representa exposición a asbestos y mesotelioma, con posible confusión por exposición ocupacional mixta no registrada, ilustra la incertidumbre por variables no medidas (Greenland et al., 1999; VanderWeele & Shpitser, 2011):

1) **Problema:** Exposicion_Mixta (ej: sílice, humos de soldadura) es un confusor que influye tanto en la probabilidad de trabajar con asbestos como en el riesgo de mesotelioma, pero no está registrada en la base de datos histórica.

2) **Consecuencia:** Al no ajustar por ella, la OR observada (ej: 4.5) es sesgada hacia arriba comparada con la OR "real" (ej: 3.0). El investigador sobrestima el efecto específico del asbesto.

3) **Solución (E-value):** Calculamos la fuerza mínima que debería tener el confusor no medido (asociación con exposición *y* con outcome) para eliminar el efecto observado. Si el E-value es alto (ej: >3.0), se requiere un confusor muy fuerte para explicar el resultado, lo que da robustez a la inferencia. Si es bajo (ej: 1.5), la asociación podría ser espuria por confusión residual.

Ejemplo 7.2: Variables proxy y equidad en salud laboral

Contexto: Migrantes en agricultura y enfermedad respiratoria, ajustando por "nivel educativo" vs "estatus migratorio" (ver Gráfico 14).

Código para Rstudio:

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_equidad <- dagitty('dag {
Exposicion_Agricola [exposure,pos="2,5"]
  Enfermedad_Respiratoria [outcome,pos="2,0"]
  Estatus_Migratorio [pos="0,2"]
  Nivel_Educativo [pos="4,2"]
  Edad [pos="1,4"]
  Tiempo_Pais [pos="3,4"]
  Lengua_Indigena [pos="0,4" latent]
  Acceso_Salud [pos="4,4" latent]
  Estatus_Migratorio -> Exposicion_Agricola
  Estatus_Migratorio -> Nivel_Educativo
  Estatus_Migratorio -> Acceso_Salud
  Nivel_Educativo -> Exposicion_Agricola
  Nivel_Educativo -> Acceso_Salud
  Lengua_Indigena -> Estatus_Migratorio
  Lengua_Indigena -> Nivel_Educativo
  Edad -> Exposicion_Agricola
  Edad -> Enfermedad_Respiratoria
```

```
  Tiempo_Pais -> Exposicion_Agricola
  Tiempo_Pais -> Nivel_Educativo
  Acceso_Salud -> Enfermedad_Respiratoria
  Exposicion_Agricola -> Enfermedad_Respiratoria}')
tidy_dag <- tidy_dagitty(dag_equidad)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name == "Exposicion_Agricola" ~ "Exposicion",
      name == "Enfermedad_Respiratoria" ~ "Outcome",
      name == "Estatus_Migratorio" ~ "Determinante_Estructural",
      name %in% c("Edad", "Tiempo_Pais") ~ "Confusor",
name == "Nivel_Educativo" ~ "Proxy_Problematico",
      name %in% c("Lengua_Indigena", "Acceso_Salud") ~ "Latente"),
    color = case_when(
      tipo == "Exposicion" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Determinante_Estructural" ~ "#8B0000",  # Rojo oscuro para
estructural
      tipo == "Confusor" ~ "#2E8B57",
       tipo == "Proxy_Problematico" ~ "#FF69B4",        # Rosa para proxy
problemático
tipo == "Latente" ~ "#C0C0C0"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
```

```
aes(x = x, y = y, fill = color),
size = 18, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.6, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.9, xlim = c(-0.8, 4.8), ylim = c(-0.5, 5.2)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
    title = "Figura 7.2: Variables Proxy y Equidad en Salud Laboral",
     subtitle = "Ejemplo 7.2: Migrantes en Agricultura y Enfermedad Respiratoria",
    caption = "Rojo: Determinante estructural | Rosa: Proxy problemático | Verde: Confusores | Gris: Variables latentes")
cat("=== ANALISIS DE ESTRATEGIAS DE AJUSTE ===\n")
cat("\n1. AJUSTE POR NIVEL EDUCATIVO (Problematico):\n")
cat("   - Educativo es proxy de Estatus_Migratorio\n")
cat("   - Ajustar por educacion introduce colision: Migr -> Educativo <- Lengua\n")
cat("   - Ademas, enmascara inequidades estructurales (Krieger, 2011)\n")
cat("\n2. AJUSTE POR ESTATUS MIGRATORIO (Estructuralmente valido):\n")
cat("   - Controla confusion sin abrir caminos espurios\n")
cat("   - Permite estimar efecto dentro de grupos migratorios\n")

cat("\n3. NO AJUSTAR (Estimacion de efecto total con disparidades):\n")
cat("   - Incluye efecto de migracion via peores trabajos\n")
cat("   - Relevante para politica de salud publica\n")
sets_ajuste <- adjustmentSets(dag_equidad,
exposure = "Exposicion_Agricola",
                        outcome = "Enfermedad_Respiratoria",
                        effect = "total",
                        type = "minimal")
cat("\nConjunto minimo de ajuste (segun DAG):\n")
print(sets_ajuste)
```

```
cat("\n=== SIMULACION: IMPACTO DEL AJUSTE POR PROXY ===\n")
set.seed(12345)
n <- 2500
edad <- rnorm(n, 35, 10)
lengua_indigena <- rbinom(n, 1, 0.3)  # Latente
acceso_salud <- rnorm(n, 0, 1)        # Latente
# Estatus migratorio: afectado por lengua (discriminacion estructural)
estatus_migratorio <- 0.7 * lengua_indigena + rnorm(n, 0, 1)
# Tiempo en pais: correlacionado con edad y estatus
tiempo_pais <- 0.6 * edad - 0.4 * estatus_migratorio + rnorm(n, 0, 2)

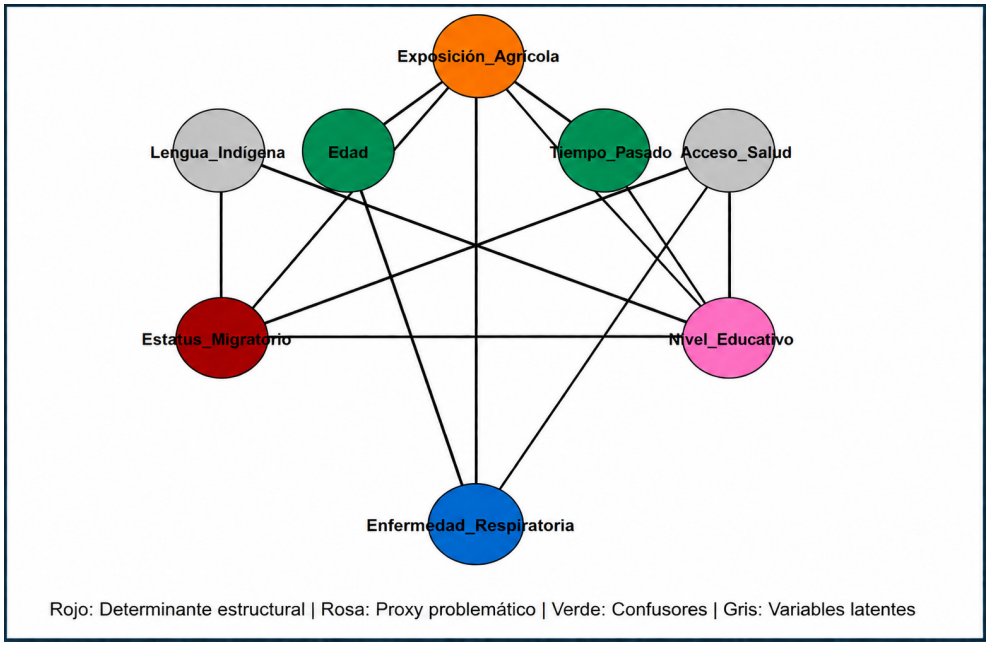
# Nivel educativo: determinado por estatus migratorio, tiempo pais y lengua
nivel_educativo <- -0.8 * estatus_migratorio + 0.3 * tiempo_pais -
                   0.5 * lengua_indigena + rnorm(n, 0, 1)
# Exposicion agricola: peores trabajos para migrantes
exposicion <- 0.6 * estatus_migratorio - 0.4 * nivel_educativo +
              0.2 * edad - 0.3 * tiempo_pais + rnorm(n, 0, 1)
# Enfermedad respiratoria
enf_resp <- rbinom(n, 1, plogis(-3 + 0.5*exposicion - 0.4*acceso_salud +
0.02*edad))
datos <- data.frame(enf_resp, exposicion, estatus_migratorio, nivel_educativo,
                    edad, tiempo_pais)
cat("\nModelo 1 - Ajuste estructuralmente valido (Estatus + Confusores):\n")
m1 <- glm(enf_resp ~ exposicion + estatus_migratorio + edad + tiempo_pais,
family = binomial, data = datos)
print(round(exp(coef(m1))["exposicion"], 2))
cat("\nModelo 2 - Ajuste por proxy (Nivel educativo en lugar de estatus):\n")
m2 <- glm(enf_resp ~ exposicion + nivel_educativo + edad + tiempo_pais,
family = binomial, data = datos)
print(round(exp(coef(m2))["exposicion"], 2))
```

```
cat("\nModelo 3 - Sin ajustar por determinante estructural (efecto to-
tal):\n")
m3 <- glm(enf_resp ~ exposicion + edad + tiempo_pais,
family = binomial, data = datos)
print(round(exp(coef(m3))["exposicion"], 2))
cat("\nConclusion: El ajuste por proxy (educacion) atenua la OR hacia la
nulidad,\n")
cat("sugiriendo falsamente que las disparidades son 'explicadas' por edu-
cacion,\n")
cat("cuando en realidad reflejan inequidad estructural (Krieger, 2011).\n")
```

Código para Rstudio generado por el autor.

Gráfico 14

DAG/Variables proxy y equidad en salud laboral



Nota. Elaboración propia. Este DAG representa un análisis de migrantes en agricultura y enfermedad respiratoria, ajustando por "nivel educativo" vs "estatus migratorio" abordándose el error ético-metodológico de ajustar por "nivel educativo" cuando la verdadera fuente de inequidad es el estatus migratorio (Krieger, 2011).

En estudios con trabajadores agrícolas, es tentador ajustar por "años de escolaridad" pensando que controlamos capacitación. Sin embargo:

- La educación es un proxy imperfecto del estatus migratorio (los migrantes indígenas tienen menos escolaridad por discriminación estructural, no por capacidad).
- Al ajustar por educación, cerramos parcialmente el camino: Migración → Educación → Exposición, pero abrimos un camino espurio vía Lengua Indígena (colisionador: Migración → Educación ← Lengua).
- Esto genera sesgo de colisión y, peor aún, enmascara la inequidad racial/étnica, sugiriendo que "la desigualdad se explica por la educación" cuando en realidad ambas son consecuencias de la discriminación estructural.

Estrategia correcta

Ajustar directamente por Estatus Migratorio (determinante estructural) permite:

1) Controlar confusión sin abrir caminos espurios.
2) Estimar efectos diferenciales por grupo (análisis estratificado).
3) No "blanquear" las disparidades sanitarias atribuyéndolas a características individuales (educación) en lugar de factores estructurales (migración).

Regla ética: Nunca ajuste por variables que sean consecuencia de la discriminación estructural si su objetivo es documentar esa misma desigualdad (Krieger, 2011).

7.4 Integración del Modelo Demandas-Recursos Laborales (JD-R) en Grafos Causales

7.4.1 El Modelo JD-R como Marco Teórico Causal

El modelo Demandas-Recursos Laborales (Job Demands-Resources, JD-R) se ha consolidado como el marco teórico dominante en la psicología de la salud ocupacional para explicar el estrés laboral y el bienestar (Bakker & Demerouti, 2007). Sin embargo, en la literatura epidemiológica tradicional, las relaciones del JD-R a menudo se reportan como asociaciones correlacionales sin una estructura causal explícita. La integración del JD-R en Grafos Acíclicos Dirigidos (DAG) permite formalizar las hipótesis sobre cómo las condiciones de trabajo influyen causalmente en la salud mental y física.

El modelo postula dos procesos psicológicos distintos:

1) **Proceso de Desgaste de Energía:** Las demandas laborales elevadas (ej. carga de trabajo, presión temporal) agotan la energía mental y física, leading al agotamiento (burnout) y problemas de salud.
2) **Proceso de Motivación:** Los recursos laborales (ej. autonomía, apoyo social) fomentan el crecimiento, el aprendizaje y el compromiso (engagement), mitigando los efectos negativos de las demandas.

7.4.2 Representación del JD-R en un DAG

Para traducir el JD-R a un DAG válido, debemos definir nodos y flechas que respeten la temporalidad y la dirección causal hipotetizada. Un DAG básico para el JD-R incluiría:

- **Nodos Exógenos:** Demandas Laborales (D), Recursos Laborales (R).
- **Nodos Mediadores:** Agotamiento/Strain (S), Compromiso/Engagement (E).

- **Nodo Desenlace:** Salud Mental/Física (H).
- **Confusores Comunes:** Factores individuales (ej. Neuroticismo, Salud basal), Factores contextuales (ej. Cultura organizacional).

La estructura causal básica se representa como:

- $D \rightarrow S \rightarrow H$ (Ruta de desgaste).
- $R \rightarrow E \rightarrow H$ (Ruta de motivación).
- $R \rightarrow H$ (Efecto directo protector de los recursos).

7.4.3 El Desafío de la Interacción (Efecto Buffer) en DAG

Una premisa central del JD-R es la hipótesis del "buffer" o amortiguación: los recursos laborales moderan (interactúan con) la relación entre las demandas y el agotamiento (Bakker & Demerouti, 2017). Es decir, el efecto causal de *D*sobre *S*depende del nivel de *R*.

En la teoría de grafos causales estándar, las interacciones estadísticas no se representan fácilmente con flechas simples. Sin embargo, para fines de identificación causal, es crucial reconocer que si existe interacción cualitativa, el efecto causal promedio (ACE) puede variar entre estratos de recursos. Según VanderWeele (2015), para representar la moderación en un marco causal, se puede utilizar un nodo de interacción explícito o estratificar el análisis. En el contexto del JD-R, si ajustamos por recursos (*R*) cuando queremos estimar el efecto total de las demandas (*D*), podríamos bloquear parcialmente el efecto si *R*actúa también como mediador en otra ruta, o podríamos estar estimando el efecto directo controlado.

Implicación para el Epidemiólogo Laboral: Si el objetivo es evaluar la eficacia de una intervención para aumentar los recursos (*R*) y así reducir el impacto de las demandas (*D*) en la salud (*H*), el DAG debe especificar que *R*no es un confusor de la relación $D \rightarrow H$, sino un modificador del efecto.

Ajustar por *R*en un modelo de regresión estándar sin términos de interacción podría ocultar el mecanismo de protección del JD-R.

7.4.4 Ejemplo Práctico: Personal Sanitario y Burnout

- **Contexto:** Hospital universitario durante periodo de alta carga.
- **Exposición (D):** Horas extras forzadas.
- **Recurso (R):** Autonomía en la toma de decisiones clínicas.
- **Desenlace (H):** Síndrome de Burnout.
- **Confusores (C):** Edad, Género, Experiencia previa.

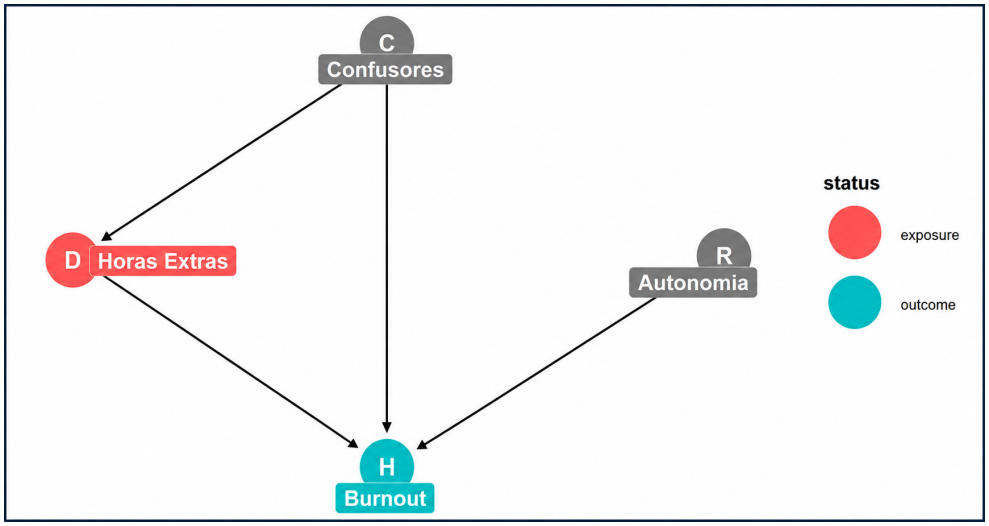
Construcción del DAG:

1) Dibujar $D \rightarrow H$.
2) Dibujar $R \rightarrow H$.
3) Dibujar $C \rightarrow D$ y $C \rightarrow H$(ej. experiencia previa reduce horas extras y reduce burnout).
4) **Hipótesis JD-R:** Incluir una flecha punteada o notar que el efecto de $D \rightarrow H$ es más fuerte cuando *R*es bajo.

Análisis de Ajuste: El conjunto de ajuste mínimo para estimar el efecto causal total de las Horas Extras (D) sobre el Burnout (H) es $\{C\}$. No se debe ajustar por *R*si se busca el efecto total de las demandas, a menos que se quiera estimar el efecto directo de las demandas manteniendo los recursos constantes (Efecto Directo Controlado). Sin embargo, para validar la hipótesis JD-R, se debe realizar un análisis de interacción $D \times R$sobre H (ver Gráfico 15).

Gráfico 15

DAG/ Modelo Causal JD-R



Nota. Elaboración propia. Análisis de un Modelo Causal JD-R con un ejemplo práctico: Personal Sanitario y Burnout

Tabla 1

Elementos de un Modelo Causal JD-R

D (Horas extras)	Nodo rojo "Horas Extras" (exposure)
H (Burnout)	Nodo turquesa "Burnout" (outcome)
R (Autonomía)	Nodo gris "Autonomía" (recurso/moderador)
C (Confusores)	Nodo gris superior "Confusores" (edad, género, experiencia)
D → H	Flecha diagonal Horas Extras → Burnout
R → H	Flecha diagonal Autonomía → Burnout
C → D	Flecha Confusores → Horas Extras
C → H	Flecha Confusores → Burnout

Nota: Elaboración propia. Elementos de un Modelo Causal JD-R. Ejemplo práctico: Personal Sanitario y Burnout.

7.4.5 Limitaciones Causales del JD-R

Es importante señalar que el JD-R original es a menudo transversal. Para que un DAG sea válido, se debe asumir la precedencia temporal: las Demandas y Recursos preceden al Agotamiento, y este precede a la Enfermedad. Estudios longitudinales son necesarios para justificar las direcciones de las flechas (Demerouti et al., 2001). Además, existe la posibilidad de causalidad inversa (workers con burnout perciben menos recursos), lo que crearía un bucle prohibido en un DAG acíclico. En tales casos, se deben usar modelos de ecuaciones estructurales o DAG con variables de retraso temporal (lagged variables).

CAPÍTULO VIII

PERSPECTIVAS FUTURAS Y CONCLUSIONES: INTEGRACIÓN DE DAG EN LA SALUD OCUPACIONAL MODERNA

8.1. Tendencias Emergentes

- **DAG dinámicos y de tiempo continuo:** Integración con modelos de procesos estocásticos para exposiciones variables en el tiempo (Keil et al., 2018).
- **Fusión con machine learning:** Uso de algoritmos de descubrimiento causal para generar hipótesis estructurales, validadas luego por expertos (Glymour et al., 2019).
- **Plataformas colaborativas:** Repositorios abiertos de DAG validados por sector industrial, facilitando meta-análisis y vigilancia comparada (Textor et al., 2016).
- **Integración con guías de práctica:** Incorporación en protocolos de la ILO, OSHA, y EU-OSHA para evaluación de riesgos basada en evidencia causal.

8.2. Recomendaciones para la Práctica

1) **Formación continua:** Incluir DAG en currículos de medicina del trabajo, higiene industrial y epidemiología.

2) **Estándares de publicación:** Exigir DAG en suplementos de revistas de salud ocupacional.

3) **Validación multidisciplinaria:** Involucrar toxicólogos, ergonomistas, economistas y trabajadores en la especificación causal.

4) **Transparencia metodológica:** Reportar conjuntos de ajuste, supuestos, y análisis de sensibilidad.

Ejemplo 8.1: DAG dinámico con medidas repetidas

- **Contexto:** Seguimiento anual de exposición a ruido y hipertensión en industria manufacturera (ver Gráfico 16).

- **Código en Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_dinamico <- dagitty('dag {
Ruido_T0 [pos="0,3"]
  Estres_T0 [pos="0,5"]
  BMI_T0 [pos="0,1"]
  Edad [pos="0,0"]
  Sexo [pos="0,6"]
  Ruido_T1 [exposure,pos="2,3"]
  Estres_T1 [pos="2,5"]
  BMI_T1 [pos="2,1"]
  Hipertension [outcome,pos="4,3"]
  Edad -> Ruido_T0
  Edad -> Hipertension
  Sexo -> Ruido_T0
```

```
  Sexo -> Hipertension
  BMI_T0 -> Ruido_T0
  BMI_T0 -> BMI_T1
  BMI_T0 -> Hipertension
  Estres_T0 -> Ruido_T0
  Estres_T0 -> Estres_T1
  Estres_T0 -> Hipertension
  Ruido_T0 -> Ruido_T1
  Ruido_T0 -> Estres_T1
  Ruido_T0 -> BMI_T1
  Ruido_T0 -> Hipertension
  Estres_T1 -> Ruido_T1
  Estres_T1 -> Hipertension
  BMI_T1 -> Ruido_T1
  BMI_T1 -> Hipertension
  Ruido_T1 -> Hipertension
  Ruido_T1 -> Estres_T1
  Ruido_T1 -> BMI_T1}')
tidy_dag <- tidy_dagitty(dag_dinamico)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name %in% c("Ruido_T0", "Ruido_T1") ~ "Exposicion",
      name == "Hipertension" ~ "Outcome",
       name %in% c("Edad", "Sexo", "BMI_T0", "Estres_T0", "BMI_T1", "Es-
tres_T1") ~ "Confusor"),
    color = case_when(
      tipo == "Exposicion" ~ "#D2691E",
      tipo == "Outcome" ~ "#0066CC",
      tipo == "Confusor" ~ "#2E8B57"),
    tiempo = case_when(
      grepl("T0", name) | name %in% c("Edad", "Sexo") ~ "T0",
      grepl("T1", name) ~ "T1",
      TRUE ~ "Outcome"))
edges <- tidy_dag$data %>%
```

```
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 19, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.7, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.8, xlim = c(-0.5, 4.5), ylim = c(-0.5, 6.5)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
    title = "Figura 8.1: DAG Dinamico con Medidas Repetidas",
    subtitle = "Ejemplo 8.1: Ruido e Hipertension en Seguimiento Anual",
    caption = "Izquierda: Baseline (T0) | Centro: Exposicion T1 | Derecha:
Outcome | Verde: Confusores time-varying")
cat("=== ESTRUCTURA TEMPORAL DEL DAG ===\n")
cat("T0 (Linea base): Confusores basales (BMI, Estres) afectan Ruido ini-
cial\n")
cat("T1 (Ano 1): Ruido acumulado afecta confusores (BMI, Estres) y out-
come\n")
cat("Problema: Estres_T1 y BMI_T1 son confusores time-varying afecta-
dos por exposicion previa\n")
cat("\n=== CONJUNTO DE AJUSTE ===\n")
sets_ajuste <- adjustmentSets(dag_dinamico,
                              exposure = "Ruido_T1",
                              outcome = "Hipertension",
```

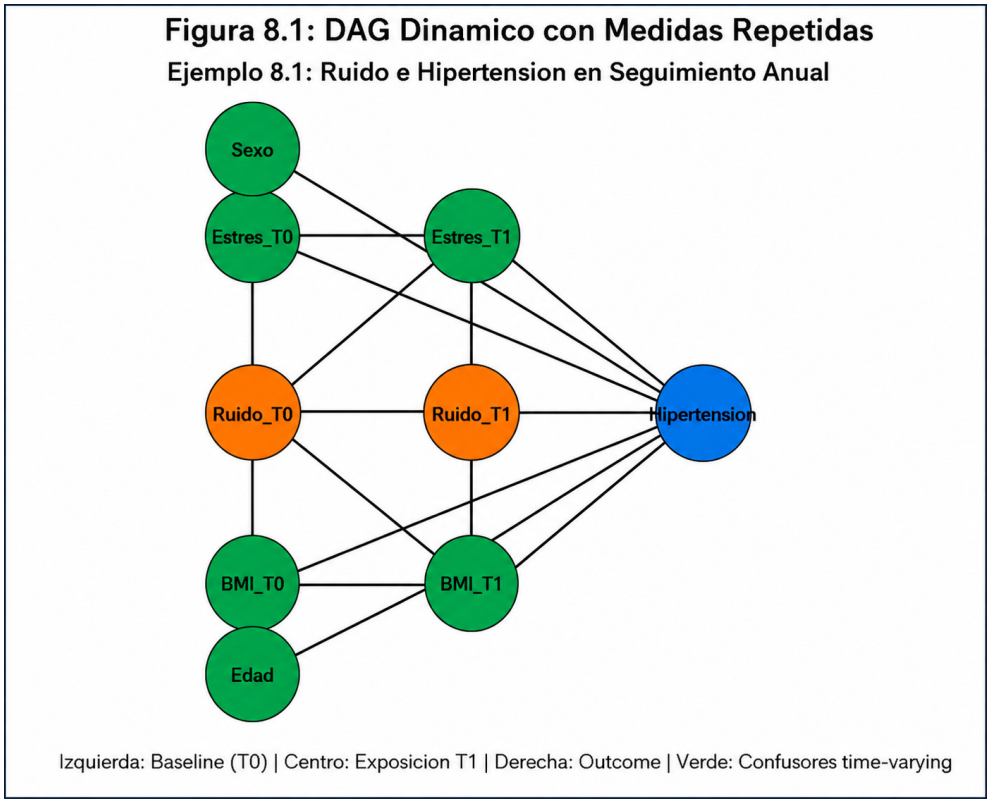
```
                    effect = "total")
print(sets_ajuste)
cat("\nNota: Ajustar por Estres_T1 o BMI_T1 (post-baseline) bloquea parte del efecto\n")
cat("causal porque son mediadores (Ruido_T0 -> Estres_T1 -> Hipertension).\n")
cat("Solucion: Usar g-computation o modelos marginales estructurales (MSM).\n")
cat("\n=== SIMULACION LONGITUDINAL ===\n")
set.seed(12345)
n <- 1500

edad <- rnorm(n, 40, 10)
sexo <- rbinom(n, 1, 0.5)
bmi_t0 <- 25 + 0.1*edad - 0.5*sexo + rnorm(n, 0, 3)
estres_t0 <- rnorm(n, 50, 10)
ruido_t0 <- 0.3*edad - 0.4*sexo + 0.1*bmi_t0 + 0.02*estres_t0 + rnorm(n, 0, 2)
bmi_t1 <- bmi_t0 + 0.2*ruido_t0 + rnorm(n, 0, 1)
estres_t1 <- estres_t0 + 0.3*ruido_t0 + rnorm(n, 0, 5)
ruido_t1 <- ruido_t0 + 0.5*estres_t1 - 0.3*bmi_t1 + rnorm(n, 0, 1)
hipertension <- rbinom(n, 1, plogis(-4 + 0.3*ruido_t0 + 0.4*ruido_t1 +
0.05*bmi_t1 + 0.02*estres_t1 + 0.05*edad))
cat("Prevalencia HTA al año 1:", round(mean(hipertension)*100,1),
"%\n")
cat("\nModelo NAIVE (ajusta por confusor T1 afectado por exposicion T0):\n")
m_naive <- glm(hipertension ~ ruido_t1 + bmi_t1 + estres_t1 + edad +
sexo,
family = binomial, data = data.frame(hipertension, ruido_t1, bmi_t1, estres_t1, edad, sexo))
print(round(exp(coef(m_naive)["ruido_t1"]), 2))
cat("\nModelo CORRECTO (ajusta solo por basales):\n")
m_correcto <- glm(hipertension ~ ruido_t1 + ruido_t0 + bmi_t0 + estres_t0 + edad + sexo,
                  family = binomial, data = data.frame(hipertension, ruido_t1,
```

ruido_t0, bmi_t0, estres_t0, edad, sexo))
print(round(exp(coef(m_correcto)["ruido_t1"]), 2))
Código para Rstudio generado por el autor

Gráfico 16

DAG/Dinámico con medidas repetidas



Nota. Elaboración propia. Este DAG es una representación de un seguimiento anual de exposición a ruido y hipertensión en industria manufacturera, representando datos longitudinales (T0 y T1), donde las variables cambian en el tiempo.

BMI_T1 y Estrés_T1 juegan un doble papel:

1) Son mediadores (Ruido_T0 aumenta el estrés y el peso).

2) Son confusores (afectan quién sigue expuesto en T1 y quién desarrolla hipertensión).

Si se ajusta por ellos en T1 (modelo naive), bloquea parte del efecto causal de la exposición previa, subestimando el daño real (Keil et al., 2018).

No ajustar por variables post-baseline que sean consecuencia de la exposición previa. Usar modelos marginales estructurales (MSM) con ponderación inversa de probabilidad o g-computation, que respetan la estructura temporal sin cortar caminos causales.

Regla:"No ajustar por el futuro para explicar el pasado; en paneles, solo ajuste por basales o use métodos dinámicos."

Ejemplo 8.2: Fusión causal para revisiones sistemáticas

- **Contexto:** Metaanálisis sobre exposición a turnos y diabetes tipo 2, combinando estudios con DAG heterogéneos (ver Gráfico 17).
- **Código para Rstudio:**

```
library(dagitty)
library(ggdag)
library(ggplot2)
library(dplyr)
dag_fusion <- dagitty('dag {
Turnos [exposure,pos="0,3"]
  Diabetes_T2 [outcome,pos="4,3"]
  Edad [pos="0,1"]
  Sexo [pos="0,5"]
  SES [pos="2,0"]
  Actividad_Fisica [pos="2,5"]
  Dieta [pos="2,1"]
  Obesidad [pos="3,3"]
  Genetica [pos="1,3" latent]
```

```
  Estudio_1 [pos="4,0"]
  Estudio_2 [pos="4,6"]
  Edad -> Turnos
  Edad -> Diabetes_T2
  Sexo -> Turnos
  Sexo -> Diabetes_T2
  SES -> Turnos
  SES -> Dieta
  SES -> Actividad_Fisica
  SES -> Diabetes_T2
  Dieta -> Obesidad
  Dieta -> Diabetes_T2
  Actividad_Fisica -> Obesidad
  Actividad_Fisica -> Diabetes_T2
  Genetica -> Turnos
  Genetica -> Obesidad
  Genetica -> Diabetes_T2
  Turnos -> Obesidad
  Turnos -> Diabetes_T2
  Obesidad -> Diabetes_T2}')
tidy_dag <- tidy_dagitty(dag_fusion)
nodes <- tidy_dag$data %>%
  distinct(name, x, y) %>%
  mutate(
    tipo = case_when(
      name == "Turnos" ~ "Exposicion",
      name == "Diabetes_T2" ~ "Outcome",
      name %in% c("Edad", "Sexo", "SES") ~ "Confusor_Core",
name %in% c("Actividad_Fisica", "Dieta") ~ "Confusor_Especifico",
      name == "Obesidad" ~ "Mediador",
      name == "Genetica" ~ "Latente",
      name %in% c("Estudio_1", "Estudio_2") ~ "Indicador_Estudio"),
    color = case_when(
      tipo == "Exposicion" ~ "#D2691E",
```

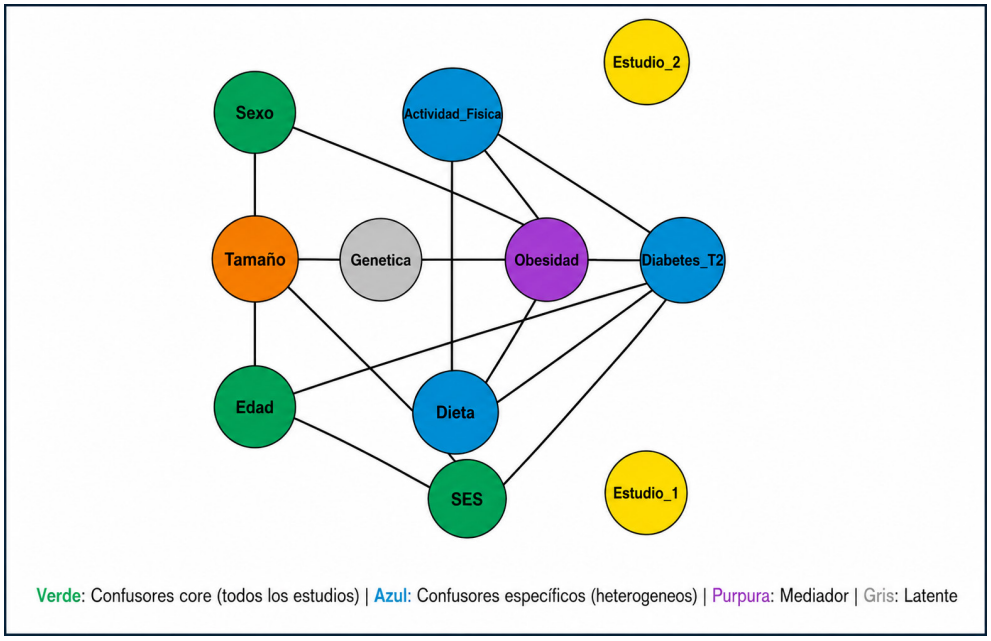
```
    tipo == "Outcome" ~ "#0066CC",
    tipo == "Confusor_Core" ~ "#2E8B57",
    tipo == "Confusor_Especifico" ~ "#4682B4",
    tipo == "Mediador" ~ "#9932CC",
    tipo == "Latente" ~ "#C0C0C0",
    tipo == "Indicador_Estudio" ~ "#FFD700"))
edges <- tidy_dag$data %>%
  filter(!is.na(to)) %>%
  select(from = name, to, x, y) %>%
  left_join(
    nodes %>% select(name, x, y) %>% rename(xend = x, yend = y),
    by = c("to" = "name"))
ggplot() +
  geom_segment(data = edges,
               aes(x = x, y = y, xend = xend, yend = yend),
               arrow = arrow(length = unit(8, "pt"), type = "closed"),
               size = 0.6) +
  geom_point(data = nodes,
aes(x = x, y = y, fill = color),
size = 18, shape = 21, color = "black", stroke = 0.8) +
  geom_text(data = nodes,
            aes(x = x, y = y, label = name),
            size = 2.6, fontface = "bold", color = "black") +
  scale_fill_identity() +
  coord_fixed(ratio = 0.9, xlim = c(-0.8, 4.8), ylim = c(-0.8, 6.5)) +
  theme_void() +
  theme(plot.margin = margin(20, 20, 20, 20)) +
labs(
    title = "Figura 8.2: Fusion Causal para Meta-Analisis",
    subtitle = "Ejemplo 8.2: Turnos de Trabajo y Diabetes Tipo 2 (Estudios
Heterogeneos)",
    caption = "Verde: Confusores core (todos los estudios) | Azul: Confuso-
res especificos (heterogeneos) | Purpura: Mediador | Gris: Latente")
cat("=== CONCEPTO DE FUSION CAUSAL ===\n")
```

```
cat("Diferentes estudios ajustan por diferentes conjuntos:\n")
cat("- Estudio 1 (cohorte nordica): Ajusta por Edad, Sexo, SES, Actividad_Fisica\n")
cat("- Estudio 2 (cohorte asiatica): Ajusta por Edad, Sexo, SES, Dieta\n")
cat("- Estudio 3 (transversal EE.UU.): Ajusta por Edad, Sexo, Obesidad (ERROR: mediador)\n")
sets_ajuste <- adjustmentSets(dag_fusion,
                              exposure = "Turnos",
                              outcome = "Diabetes_T2",
                              effect = "total")
cat("\nConjunto suficiente valido en todos los estudios (backdoor criterion):\n")
print(sets_ajuste)
cat("\n=== SIMULACION DE HETEROGENEIDAD ENTRE ESTUDIOS ===\n")
set.seed(12345)
simular_estudio <- function(n, nombre, ajusta_dieta = FALSE,
                            ajusta_actividad = FALSE, ajusta_obesidad = FALSE) {
edad <- rnorm(n, 45, 10)
  sexo <- rbinom(n, 1, 0.4)
  ses <- rnorm(n, 0, 1)
  genetica <- rnorm(n, 0, 1)
turnos <- rbinom(n, 1, plogis(-1 + 0.02*edad - 0.3*sexo - 0.4*ses + 0.3*genetica))
  dieta <- 0.5*ses + rnorm(n, 0, 1)
  actividad <- -0.4*ses + rnorm(n, 0, 1)
 obesidad <- 0.3*turnos + 0.4*dieta - 0.3*actividad + 0.5*genetica + rnorm(n, 0, 1)
  logit_diab <- -4 + 0.4*turnos + 0.05*edad + 0.2*sexo + 0.3*ses +
                0.2*dieta - 0.15*actividad + 0.3*obesidad + 0.4*genetica
  diabetes <- rbinom(n, 1, plogis(logit_diab))
   datos <- data.frame(diabetes, turnos, edad, sexo, ses, dieta, actividad, obesidad)
  # Ajuste según lo que mide cada estudio
```

```
  vars_ajuste <- c("edad", "sexo", "ses")
  if(ajusta_dieta) vars_ajuste <- c(vars_ajuste, "dieta")
  if(ajusta_actividad) vars_ajuste <- c(vars_ajuste, "actividad")
  if(ajusta_obesidad) vars_ajuste <- c(vars_ajuste, "obesidad")
  formula <- as.formula(paste("diabetes ~ turnos +", paste(vars_ajuste, co-
llapse = " + ")))
  modelo <- glm(formula, family = binomial, data = datos)
  or <- exp(coef(modelo)["turnos"])
  ci <- exp(confint(modelo)["turnos", ])
  cat(nombre, "(n=", n, "): OR =", round(or, 2),
"(95% CI:", round(ci[1], 2), "-", round(ci[2], 2), ")\n")
  return(c(or = or, lower = ci[1], upper = ci[2], n = n))}
est1 <- simular_estudio(800, "Estudio 1 (Nordico)", ajusta_actividad =
TRUE)
est2 <- simular_estudio(1200, "Estudio 2 (Asiatico)", ajusta_dieta = TRUE)
est3 <- simular_estudio(600, "Estudio 3 (EE.UU.)", ajusta_obesidad =
TRUE)
cat("\n=== META-ANALISIS (INVERSE VARIANCE) ===\n")
# Ponderacion simple por tamano (simulacion)
pesos <- c(est1["n"], est2["n"], est3["n"]) / sum(c(est1["n"], est2["n"], es-
t3["n"]))
or_pooled <- sum(c(est1["or"], est2["or"], est3["or"]) * pesos)
cat("OR Ponderada (naive):", round(or_pooled, 2), "\n")
cat("\nNota: El Estudio 3 sesga el meta-analisis al ajustar por obesidad
(mediador).\n")
cat("Fusion causal valida requiere: (1) Homogeneidad de efecto causal,
o\n")
cat("(2) Estratificacion por conjunto de ajuste (VanderWeele et al.,
2020).\n")
```

Gráfico 17

DAG/ Fusión causal para revisiones sistemáticas



Nota. Elaboración propia. Este DAG representa un metaanálisis sobre exposición a turnos y diabetes tipo 2, combinando estudios con DAG heterogéneos y abordando la heterogeneidad metodológica entre estudios en revisiones sistemáticas (VanderWeele et al., 2020).

El problema de los "ajustes variables"

En metaanálisis sobre turnos y diabetes:

- Estudio 1 ajusta por actividad física (azul).
- Estudio 2 ajusta por dieta (azul).
- Estudio 3 ajusta por obesidad (púrpura) —esto es un error: la obesidad es mediador (turnos alteran el ritmo circadiano → obesidad → diabetes), no confusor.

Al combinar estos estudios mediante metaanálisis convencional (inversa de varianza), el Estudio 3 sesga el efecto pooled hacia la nulidad (subestima el daño real de los turnos) porque ajusta por una consecuencia intermedia de la exposición.

Solución: Fusión causal válida

No todos los estudios son combinables matemáticamente equivalentes. Se requiere:

1) Estratificación por conjunto de ajuste: metaanálisis separados para estudios que ajustan por el mismo backdoor set válido.
2) Análisis de sensibilidad: Evaluar cómo cambia el efecto pooled al excluir estudios con ajustes problemáticos (mediadores).

Regla: Nunca se debe meta-analizar estudios que ajusten por mediadores junto con estudios que no lo hagan, salvo que use métodos de mediación causal para descomponer efectos directos e indirectos.

8.3 Comparación Metodológica entre DAG y Modelos de Ecuaciones Estructurales (SEM)

8.3.1 Contexto: El Dominio del SEM en Psicología de la Salud Ocupacional

Los Modelos de Ecuaciones Estructurales (SEM) han sido durante décadas la metodología predominante en psicología de la salud ocupacional para testar modelos teóricos como el JD-R, el Modelo Demandas-Control-Apoyo, y el Modelo de Desequilibrio Esfuerzo-Recompensa (Kline, 2015). Sin embargo, los DAG representan un enfoque complementario (y en algunos aspectos, superior) para la inferencia causal. Esta sección compara ambas metodologías para guiar al investigador en su selección apropiada.

8.3.2 Diferencias Fundamentales entre DAG y SEM

Tabla 2

Diferencias Fundamentales entre DAG y SEM

Característica	DAG	SEM
Propósito principal	Identificación causal y control de sesgos	Test de ajuste del modelo y estimación de efectos
Suposiciones causales	Explícitas y visuales antes del análisis	Implícitas en la especificación del modelo
Variables latentes	Limitadas (requieren extensiones)	Manejo nativo con indicadores múltiples
Causalidad inversa	No permitida (acíclico por definición)	Permitida en modelos no recursivos
Criterio de evaluación	Identificabilidad (back-door criterion)	Índices de ajuste (CFI, RMSEA, SRMR)
Software común	DAGitty, R (ggdag)	Mplus, AMOS, lavaan (R)

Fuente: Adaptado de Pearl (2009) y Kline (2015)

8.3.3 Fortalezas de los DAG sobre SEM en Epidemiología Ocupacional

Transparencia de Suposiciones Causales: Los DAG obligan al investigador a declarar explícitamente todas las suposiciones causales antes de analizar los datos. En SEM, es común especificar modelos basados en la significancia estadística más que en teoría causal sólida (Greenland et al., 1999).

Identificación de Sesgos de Selección: Los DAG son superiores para identificar sesgos de selección (como el Efecto del Trabajador Sano) mediante la detección de colisionadores. El SEM tradicional no alerta automáticamente sobre estos sesgos (Hernán & Robins, 2020).

Conjuntos Mínimos de Ajuste: Los DAG permiten identificar algorítmicamente el conjunto mínimo de variables de ajuste, evitando el sobreajuste. El SEM típicamente incluye todas las variables disponibles, lo que puede introducir sesgo (Textor et al., 2016).

Claridad en Mediación vs. Confusión: Los DAG distinguen claramente entre variables que deben ajustarse (confusores) y variables que no deben ajustarse (mediadores, colisionadores). En SEM, esta distinción es menos evidente (VanderWeele, 2015).

CAPÍTULO IX

INFERENCIA CAUSAL EN LA EVALUACIÓN DE TERAPIAS DE LARGA LATENCIA: UNA FORMALIZACIÓN TEÓRICA DEL SESGO DE TIEMPO DE INMORTALIDAD MEDIANTE GRAFOS ACÍCLICOS DIRIGIDOS

9.1. Introducción

Rodríguez-Villamizar (2017) menciona que, en el campo de la medicina y la epidemiología tradicional, una causa se relaciona con una exposición, condición, acontecimiento o característica que antecede en un determinado momento por su presencia, y puede influir en la probabilidad de que un evento ocurra, considerando igualmente condiciones fijas alrededor del mismo, como lo es el inicio de una enfermedad. Profundizando en este pensamiento, la epidemiología laboral se enfoca en el proceso metódico mediante el cual se determina si una relación o asociación estadística entre dos o más condiciones representa verdaderamente una relación

causa-efecto o es una simple correlación. A este proceso que se adentra a demostrar una conexión etiológica (comprobable) entre una exposición (virus/fármaco) y un desenlace (enfermedad), se le conoce como inferencia causal.

Schoenbach (2004) alude que, mientras la inferencia estadística saca conclusiones generales sobre una población con base en datos extraídos de una muestra, encaminándose a establecer correlaciones y prediciendo patrones, la inferencia causal busca determinar con exactitud las diferentes variables que intervienen y modifican el resultado producido. Esta fortaleza metódica de la inferencia causal ayuda en la epidemiología laboral a las exigencias observacionales y la certificación de validez interna y externa de sus estructuras causales, tanto en el abordaje de la prevención de padecimientos de salud en los trabajadores como en las lesiones ocupacionales resultantes (Benavides, 2021).

Para Bacallao (2012), en la investigación biomédica actual la concepción de causalidad y sesgo estánunidas dentro del análisis evaluativo de diversas condiciones de exposición que intervienen en un desenlace o resultado producido, entendiendo al respecto que los sesgos se definen como errores sistemáticos que interfieren en la claridad y veracidad relacional de las variables de exposición, indicando erróneamente que una variable es causante de otra. Tal es el caso del sesgo de tiempo de inmortalidad, el cual es un error sistemático muy recurrente en los estudios observacionales (p. ej., fármacos para enfermedades crónicas o tratamientos oncológicos), ya que surge ante la existencia de un periodo de seguimiento en pacientes que han ingresado en un estudio médico y se tiene que determinar la relación de la variable de exposición, el tiempo de seguimiento y el evento resultante, el cual no ocurre por definición (Pérez de Arenaza y Pizarro, 2021).

El sesgo de tiempo de inmortalidad, en el campo biomédico, con frecuencia aparece cuando una cohorte de pacientes expuestos debe sobrevivir para ser intervenida o tratada *a priori*, generando una incidencia menor espuria del resultado en ellos al compararse con otro grupo o cohorte que

no ha sido expuesto; este sesgo falsea los hallazgos y los dirige hacia una aparente mayor efectividad o supervivencia. En otras palabras, la falsedad o distorsión aparece porque la asignación del tratamiento ocurre después de que el paciente ha entrado en la cohorte o grupo de observación; es decir, el cronómetro observador comienza desde el ingreso y no desde la exposición real, contándose al paciente como superviviente en la cohorte expuesta en un periodo (erróneo) cuyo resultado lo califica para estar allí. Al no corregirse, este sesgo acarrea resultados, falsedades, beneficios o intervencioneserróneas, así como la toma de decisiones desinformadas (Pérez de Arenaza y Pizarro, 2021).

Fox et al. (2024) apuntan que, para disminuir los sesgos y aseverar la validez de la inferencia causal, las herramientas y metodologías de frecuente uso son los DAG, ya que modelan y mapean visualmente las estructuras causales entre variables, determinando las posibles líneas direccionales causales y aquellas que se deben bloquear. Los DAG son instrumentos que permiten identificar los confusores y cuáles de las variables se deben ajustar, lo cual clarifica la causalidad, guiando y previniendo cualquier problema que se pueda presentaren el diseño temporal y la definición del estudio base observacional.

9.2. Aspectos metodológicos

9.2.1. Objetivos de la investigación

La epidemiología laboral continuamente requiere optimizar los diseños de estudios, sobre todo aquellos observacionales, por la presencia de sesgos como el de inmortalidad, que socavan la validez de los resultados de ciertas investigaciones. Igualmente, esta disciplina debe perseguir perfeccionar la precisión de sus análisis de datos, con lo cual se fortifican los procesos de toma de decisiones. Por tal razón, el objetivo general de esta investigación es: desarrollar un manual y guía académico-práctica con base en evidencia científica y metodológica sobre el uso de los DAG en la epidemiología laboral.

Para ello, los objetivos específicos investigativos seránlos siguientes:

- Definir los fundamentos teóricos de la teoría causal y su representación gráfica a través del uso de los DAG en el ámbito de la salud ocupacional y su subdisciplina científica, la epidemiología laboral.
- Analizarlas diferentes apreciaciones metodológicas para identificar y controlar la aparición de sesgos (confusores, colisionadores, sesgo de selección y sesgo de inmortalidad) en contextos de exposición y resultados de estudios ocupacionales u observacionales.
- Determinar directrices metodológicas para el diseño de estudios, el análisis causal, la implementación de software especializado, la validación y la comunicación eficaz de resultados basados en DAG dentro de entornos laborales.
- Interpretar las consideraciones éticas, críticas y limitaciones contemporáneas con respecto a la utilidad de los DAG en la investigación de la salud ocupacional y su subdisciplina científica, la epidemiología laboral.
- Aplicarteóricamente, en el marco de la inferencia causal, la modelización de grafos en estudios observacionales con presencia de sesgos epidemiológicos, como el de tiempo de inmortalidad, en la evaluación de terapias de larga latencia.

9.2.2. Metodología aplicada

- **Enfoque de la Investigación**

La investigación posee un enfoque cualitativo-analítico y de carácter teórico-conceptual. Un enfoque analítico compila datos informativos no numéricos de un fenómeno o evento (estudio de casos) para su análisis, comprensión e interpretación, validando externamente sus estudios mediante trabajos de campo como: técnicas de observación, análisis de significados, textos, discusiones, entrevistas y estudios estructurales, con los cuales se otorga fundamento teórico-conceptual a las investigaciones y posibilidad

metodológica de aplicación. En el caso de esta investigación, el enfoque es la fundamentación en la modelación formal y el análisis estructural de la causalidad, mediante el uso de grafos matemáticos que puedan darle lógica a las relaciones de dependencia epidemiológica en el campo de la salud ocupacional (Baena, 2017; Hernández-Sampieri y Mendoza, 2018).

- **Tipo de Investigación**

La tipología investigativa es de tipo documental e instrumental. Documental, por la indagación sistemática y metodológica de documentación física y digital que proporciona específicamente datos informativos científicos que sustentan la investigación en curso e; instrumental, ya que con el objetivo de construir nuevos conocimientos, disipar dudas y resolver problemas —en este caso, formalizar metodológicamente los DAG con el fin de resolver problemas de validez interna en estudios de salud ocupacional—, se aplican esquemas minuciosos de recopilación, organización y análisis de información previa, que permitan darle sustento y profundidad al estudio del tema concreto (Baena, 2017; Hernández-Sampieri y Mendoza, 2018).

- **Diseño de la Investigación** Esta investigación es no experimental, ya que estratégicamente se lleva a cabo una revisión sistemática de conceptos, fundamentos, cuerpos teóricos, formalización de modelos y marco metodológico sin manipular de forma deliberada las variables del caso de estudio; es decir, se aborda el fenómeno en su estado natural, tal como ocurren los hechos, para explicarlos, describirlos y correlacionarlos. Por otro lado, la investigación se realiza mediante una indagación de material documental impreso y digital de fuentes secundarias (Baena, 2017; Hernández-Sampieri y Mendoza, 2018).
- **Estrategias de Recolección de Datos** La estrategia de recolección de información es mediante el método de revisión, análisis, identificación y selección documental, abarcando la literatura científica arbitrada y registrada con orden (indexada), entre la que se cuenta: estudios de casos, guías metodológicas, ensayos, tesis, libros, artícu-

los científicos y de revisión, manuales técnicos de inferencia causal y otras evidencias textuales disponibles de la temática en estudio. Con respecto a la elaboración de los DAG, estos se realizaron a partir de la información extraída de la documentación sobre exposiciones ocupacionales y sesgos clásicos en medicina ocupacional y su subdisciplina, la epidemiología laboral (Baena, 2017; Hernández-Sampieri y Mendoza, 2018).

- **Técnica de Análisis de Datos** El proceso de análisis de información será el análisis de contenido, el cual será organizado jerárquicamente según la secuencia de los capítulos y los objetivos planteados, partiendo del contenido que fundamenta la teoría, los de utilidad para realizar los grafos, así como los contenidos escogidos para la formalización teórica. El análisis de contenido en esta investigación se fundamenta en precisar textos, conceptos, hallazgos y cualquier otra información con la cual realizar inferencias válidas y actualizadas sobre el tema (Baena, 2017; Hernández-Sampieri y Mendoza, 2018).

9.3. Resultados

9.3.1. Ontología del tiempo en el DAG

La ontología del tiempo, filosóficamente hablando, es el estudio de la naturaleza del tiempo: ¿qué es en su esencia más pura?, ¿qué es el tiempo verdaderamente? Velasco-Cobos (2024), con respecto a esto, menciona que el tiempo existe como una intuición humana, con dificultad para ser explicado, pero concebido desde la capacidad de desplazamiento en el espacio, siendo una linealidad bidireccional de pasado, presente y futuro; es decir, es una concepción que arraiga una creencia de movimiento de algo físico cuyo camino siempre es lineal, como una flecha que se direcciona en un solo sentido.

La conceptualización del tiempo se presenta como una estampa lineal de que un evento sucede dentro de otro, siendo ese otro el tiempo: algo que existe, es real y valedero, que proporciona conocimiento, otorga espa-

cio a la materia (cosas), las lleva y trae, provocando por definición que el tiempo sea un algo relacional. Se conceptualiza como tiempo la facultad de probabilidad de que un objeto/materia se relacione con otros y con el espacio que lo contiene, y origine una existencia que, al ser analizada, incline una decisión. En definitiva, el tiempo es más que un contenedor donde acontecen sucesos, no existe por sí mismo; es la relación existente entre un evento y otro, es el cambio y la correlación entre las cosas (Velasco-Cobos, 2024).

Esta perspectiva ontológica del tiempo, referida por Velasco-Cobos (2024), al ser aplicada al contexto de los DAG, transforma al tiempo de materia u objeto, redefiniéndolo a cualidad o atributo relacional, donde eventos, variables o tareas en una línea bidireccional e irreversible se relacionan secuencialmente e implantan una estructura alineada a la causalidad. La ontología del tiempo en los DAG se manifiesta al estudiarse la naturaleza, estructura y orden estricto de los acontecimientos; además, por ser estructuras matemáticas que se construyen con nodos y aristas (flechas), no existe la presencia de nudos o bucles que originen retroalimentación, persistiendo en un camino lineal, progresivo, modelado por un flujo temporal que conlleva a la causalidad (Rodríguez-García y Hoehndorf, 2018; Gomstyn y Jonker, 2024).

9.3.2. Mecanismos de selección dinámica

En la epidemiología laboral, la metodología de la inferencia causal se enfoca en identificar las relaciones de causa y efecto entre los factores de riesgo y las enfermedades ocupacionales. Los hallazgos de esta investigación precisan que los mecanismos de selección dinámica determinan el valor causal explicativo de la selección de las variables, con lo cual se diseña el análisis, se ajusta el modelo y se evitan los sesgos. En este contexto, los DAG, como métodos gráficos y matemáticos, esclarecen las posibles asociaciones causales, las asociaciones en rutas no causales o las propias de un sesgo (Sánchez-Niubò et al., 2016).

Los mecanismos de selección permiten, por ejemplo, ajustar una variable que es comúnmente consecuencia/efecto de otras dos variables (sesgo de selección), abriendo rutas no causales entre ellas que en principio no estaban asociadas, lo cual genera correlaciones espurias; por otro lado, los mecanismos de selección ayudan a bloquear rutas causales entre exposición y efecto (resultado). Resulta, entonces, que dichos mecanismos de selección de variables son decisivos para precisar efectos causales sin incluir sesgos y, en concordancia con el uso de los DAG, se convierten en una parte indispensable y habitual de la investigación epidemiológica laboral (Velarde-Martínez, 2018).

9.3.3. Soluciones estructurales

Los hallazgos investigativos refieren que las soluciones estructurales en el contexto de los DAG son de gran utilidad en el campo de la epidemiología laboral en concordancia con la informática, ya que, por su arquitectura de red (software o programas estructurados), ayudan en la detección de ciclos y en la clasificación topológica. En cuanto a la detección de ciclos, las soluciones estructuradas permiten, en el diseño del análisis de los DAG, verificar que en un recorrido los vértices en estado "en proceso" no presenten, por fundamento, ningún ciclo ni bucle en los grafos, ya que si se llegase a detectar alguno, la estructura sencillamente deja de ser un DAG. Con respecto a la clasificación topológica, siendo un proceso de importancia en los análisis de los DAG, las soluciones estructuradas contribuyen a organizar con rigurosidad secuencial los vértices de forma lineal/bidireccional, donde cada nodo precede a sus sucesores (Gomstyn y Jonker, 2024).

Las soluciones estructuradas, al contribuir con todo lo anterior, también colaboran con la programación de la ruta crítica de proyectos (el paso a paso de las actividades necesarias para culminar a tiempo), la planificación de las tareas (estructuras representativas de operaciones dependientes con el fin de optimizar el paralelismo) y, en sistemas distribuidos y control de versiones, gestionan las relaciones de dependencia (Gomstyn y Jonker, 2024). Un ejemplo de esas soluciones estructurales es la plataforma web

y paquete de R *DAGitty*, que posee una especial utilidad en el campo de la inferencia causal de los estudios de la epidemiología laboral, al permitir dibujar DAG, identificar conjuntos de ajuste mínimos, detectar colisionadores y exportar a R/Stata (Textor et al., 2016).

9.3.4. Implicaciones en la toma de decisiones

Los DAG son estructuras de datos que, a su vez, definen tareas procesadoras de datos acíclicos que pueden articular flujos de trabajo automatizados y complejos, limitando la intuición y guiando la toma de decisiones, ya que la recopilación e interpretación de datos permite el ajuste de variables antes del diseño de análisis. Luego, en los estudios observacionales, contribuyen a mejorar la precisión de los factores de riesgo y las medidas preventivas en las enfermedades profesionales identificadas por la epidemiología laboral (Gomstyn y Jonker, 2024).

Los DAG, ante todo, reducen los sesgos y los riesgos, ya que la evidencia empírica que proporciona el procesamiento de datos fundamenta la toma de decisiones. Además, identifican patrones y tendencias por la facultad predictiva de su análisis y la automatización descriptiva. Finalmente, la solidez objetiva de su análisis y la secuencia lineal y acíclica de su metodología reducen la incertidumbre (Sánchez-Niubò et al., 2016; Mucci, 2025). Los DAG, como procesadores de datos eficientes y escalables, gestionan, procesan y descartan importantes volúmenes de información con la finalidad de que la toma de decisiones sea organizada, precisa y eficaz (Gomstyn y Jonker, 2024).

9.3.5. Sesgo de "tiempo inmortal" en la evaluación de terapias de larga latencia

Uno de los resultados más relevantes de esta investigación es la utilidad de los DAG en la determinación del sesgo de "tiempo inmortal" en la evaluación de terapias de larga latencia, evidenciando que, en este tipo de estudios observacionales, este error no debe clasificarse como un sesgo único, ya que es una consecuencia/desenlace de dos errores que se co-

meten al principio del estudio: la falta de alineación entre el momento de asignación a la exposición y los criterios de elegibilidad (Hernán et al., 2024). La aparición de este error de "tiempo inmortal" se debe a que la data informativa usada es posterior al momento de la elegibilidad para definir la exposición, y a que los criterios de elegibilidad se fundamentaron en una data errada (Yang et al., 2025).

9.4. Discusión

En cualquier estudio de inferencia causal lo relevante es identificar con eficacia aquellos factores que intervienen en la observancia causal verdadera y precisa de una asociación entre una exposición y la aparición de una enfermedad en determinada población de interés. Se debe tomar en cuenta que es ingenuo pensar que en un estudio no puedan surgir errores sistemáticos y aleatorios (sesgos) propios de la dinámica investigativa. Por tal razón, con el propósito de contribuir a socavar mínimamente la presencia de determinados sesgos, se exponen ciertas investigaciones que contribuyen a determinar la importancia del uso de los DAG como herramientas metodológicas que ayudan a comprender las fuentes de estos errores y cuándo son un impedimento en la identificación de la causalidad. Entre estas investigaciones se cuentan:

Un artículo elaborado por Fox, M., Arah, O., Swanson, S. y Viallon, V. en el 2024, titulado: *"Causal diagrams to evaluate sources of bias"* (Diagramas causales para evaluar las fuentes de sesgo). El objetivo de esta investigación es destacar la importancia del uso metodológico de los diagramas causales en los estudios de epidemiología —en este caso los DAG— como herramientas que contribuyen a identificar, comprender y evaluar las fuentes de sesgos o errores sistemáticos en estudios observacionales, con base en el proceso de inferencia causal (relaciones de causa y efecto real). Permitiendo además, mejorar la comunicación interdisciplinaria en el campo científico de la salud mediante el esclarecimiento de desacuerdos en los análisis causales en cuanto a las variables que requieren ajustes (Fox et al., 2024).

Los autores se enfocan en demostrar que los sesgos se generan por estimaciones que difieren del verdadero efecto causal (error sistemático) y por razones distintas al error aleatorio, como por ejemplo el aumento de una muestra. Conociendo que los epidemiólogos otorgan relevancia a la confusión, al sesgo de información y al sesgo de selección como fuentes principales del error sistemático, la metodología aplicada en esta investigación se direcciona a formalizar teórica y gráficamente los diagramas causales, especialmente los DAG, a través de su representación en grafos y del análisis de ejemplos de exposición encontrados en la literatura de inferencia causal (Fox et al., 2024).

Los autores abordan tópicos como los DAG y las fuentes específicas de sesgo, los DAG y las fuentes múltiples de sesgo, los DAG con signos, y el uso de los DAG en la síntesis de la evidencia. Esto se realiza con el propósito de respetar el principio de temporalidad de la causalidad, las relaciones temporales y la indexación de variables, así como determinar las rutas dirigidas y no dirigidas entre las variables y monitorear la generación de datos, entre otros aspectos, para conocer las fuentes de los sesgos y cómo interfieren en la identificación de la causalidad (Fox et al., 2024).

Las conclusiones de esta investigación refieren que los DAG son útiles por tres razones principales: Primero, porque hacen más explícitas las hipótesis, identificando entre ellas puntos de desacuerdo y contribuyendo a detectar fuentes relevantes de sesgos, tanto en estudios individuales como en grupos de evidencias; segundo, porque al complementarse con otros métodos motivan el análisis de sesgo o de sensibilidad, con lo cual se identifican fuentes de sesgos que influyen en las conclusiones sobre peligros concretos; y finalmente, porque facilitan una estructura sistemática de identificación de riesgos, la comunicación de modelos de trabajo, la emisión de juicios de calidad de estudios subyacentes, y la presentación de la evidencia más útil y eficiente (Fox et al., 2024).

Por otra parte, se cuenta con la investigación realizada por Hernán, M., Sterne, J., Higgins, J., Shrier, I. y Hernández-Díaz, S. en el 2024, intitula-

da: *"A structural description of biases that generate immortal time"* (Una descripción estructural de los sesgos que generan el «tiempo inmortal»). Para estos autores, en el campo de la salud, el tiempo de inmortalidad se origina cuando un grupo de pacientes que serán parte de un estudio observacional son seleccionados en función de ciertos criterios de elegibilidad posteriores a la exposición o al tratamiento asignado. En otras palabras, el sesgo de tiempo inmortal es un lapso que se incluye en el análisis observacional durante el cual el paciente, por definición, no desarrolló el resultado o evento esperado; es decir, si el resultado era la muerte y esta no ocurrió, la persona es «inmortal». Este sesgo se ha presentado durante dos siglos en los estudios observacionales de supervivencia, arrojando estimaciones sesgadas de los verdaderos efectos (Hernán et al., 2024).

Por este motivo, el objetivo de esta investigación es revisar y demostrar la fuente del origen del sesgo de inmortalidad, proponiendo metodológicamente la utilización de los diagramas causales y los DAG como herramientas que desglosan la lógica y el fundamento matemático de dicho sesgo, ya que mediante la representación gráfica se resumen los estudios observacionales y los enfoques analíticos, evitándose la generación del tiempo inmortal (Hernán et al., 2024).

Esta investigación resalta que los DAG ayudan a estructurar ventanas de tiempo que permiten visualizar eventos futuros y posibles rutas causales alternas, condicionando el análisis para que, con estricto rigor, se alineen el tiempo cero (inicio del seguimiento), la variable de elegibilidad y la asignación de la exposición, variables que en su mayoría son dependientes del tiempo. Concluyen que lo que se ha denominado en los estudios observacionales como sesgo de tiempo inmortal no es un tipo de sesgo único, sino el resultado de dos errores o situaciones desacertadamente coordinadas al principio del análisis, las cuales serían: Los criterios de elegibilidad (sesgo de selección) y la equívoca clasificación en la asignación del tratamiento luego de iniciar el proceso de seguimiento u observación (Hernán et al., 2024).

Otro artículo de relevancia fue el elaborado por Yang, G., Burgess, S. y Schooling, C. en el 2025, denominado: "Illustrating the structures of bias from immortal time using directed acyclic graphs" (Ilustración de las estructuras de sesgo a partir del «tiempo inmortal» mediante grafos acíclicos dirigidos). Para estos autores, el tiempo inmortal es un lapso diseñado para observar y realizar un seguimiento en el que, por definición, no puede producirse la muerte o el resultado del estudio en los pacientes. Este sesgo ha estado presente por mucho tiempo en las observaciones científicas realizadas por la epidemiología sin que se hayan concretado sus causas, estructura y, menos aún, explicado de forma sistemática. Esta carencia es el objetivo principal de este estudio: fundamentar y explicar sistemáticamente las fuentes de origen y la estructura del sesgo de tiempo inmortal (Yang et al., 2025).

Para cimentar metodológicamente esta investigación se realizó un estudio base a manera de ilustración, con el fin de mostrar el surgimiento del tiempo inmortal y presentar la estructura del sesgo utilizando DAG que expusieron y especificaron las variables relacionadas con el tiempo. Aparte, se llevó a cabo otro análisis estructural obviando el tiempo inmortal en presencia de riesgos competitivos, analizándose cómo dichas estructuras se presentan en variados diseños de estudio en farmacoepidemiología y ofreciéndose soluciones para abordar y controlar la aparición del sesgo estudiado (Yang et al., 2025).

Los resultados de esta investigación fueron: Una falta de alineación entre el momento de asignación de la exposición y los criterios de elegibilidad como causa fundamental del origen del tiempo inmortal; por otro lado, se concretó que la aparición del sesgo se debe al uso de datos informativos posteriores a la elegibilidad para definir la exposición, o por usar datos informativos luego de la exposición para definir la elegibilidad (Yang et al., 2025).

En cuanto a la estructura del sesgo inmortal, resultó que este se deriva del sesgo de confusión por supervivencia hasta la asignación de la exposición

o, igualmente, por el sesgo de selección, ya que esta se realiza en función de la supervivencia hasta la elegibilidad. Por otro lado, el análisis estructural obviando el tiempo inmortal no resuelve los sesgos de confusión o de selección, y los riesgos competitivos presentes llegan a agravar el sesgo; sin embargo, los datos basales —como la asignación correcta de la exposición y la elegibilidad—, así como la exclusión de los pacientes con exposición previa, pueden evitar el sesgo derivado del tiempo inmortal (Yang et al., 2025).

Se concluye que la comprensión del sesgo derivado del tiempo inmortal, partiendo del sesgo de confusión o de selección, permite la identificación de las fuentes de origen y de cuándo ocurren, ayudando así a su prevención o mitigación (Yang et al., 2025).

CONCLUSIÓN

Los Grafos Acíclicos Dirigidos representan un avance metodológico sustancial en epidemiología laboral. No eliminan la incertidumbre inherente a los estudios observacionales, pero la hacen explícita, manejable y comunicable. Al forzar la especificación de supuestos causales, previenen errores de ajuste, reducen sesgos estructurales y fortalecen la validez interna de las inferencias sobre riesgos ocupacionales. Su uso responsable, combinado con rigor ético, transparencia y validación empírica, posiciona a la epidemiología laboral como una disciplina capaz de generar evidencia causal robusta para la prevención, la regulación y la protección de la salud de los trabajadores.

La adopción de DAG no es opcional en la epidemiología contemporánea; es un imperativo metodológico y ético. Solo mediante la explicitación clara de nuestras suposiciones causales podemos avanzar hacia una salud ocupacional basada en evidencia, equitativa y efectiva.

REFERENCIAS BIBLIOGRÁFICAS

Alqahtani, J. S., Alqahtani, A. M., Almutairi, A. F., & Alghamdi, M. A. (2024). The effect of cumulative night shift duties on insomnia, fatigue, and depression among nurses. Heliyon, 10(13), e33097. https://doi.org/10.1016/j.heliyon.2024.e33097

Arrighi, H. M., & Hertz-Picciotto, I. (1994). The evolving concept of the healthy worker survivor effect. *Epidemiology, 5*(2), 189–196. https://doi.org/10.1097/00001648-199403000-00009

Bacallao, J. (2012). Enfoques modernos del sesgo y la causalidad en la investigación epidemiológica. Revista Cubana de Salud Pública, 38(5),686-701. https://www.redalyc.org/articulo.oa?id=21425053003

Baena, G. (2017). Metodología de la investigación. Grupo Editorial Patria. http://www.biblioteca.cij.gob.mx/archivos/materiales_de_consulta/drogas_de_abuso/articulos/metodologia%20de%20la%20investigacion.pdf

Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, *22*(3), 309–328. https://doi.org/10.1108/02683940710733115

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273–285. https://doi.org/10.1037/ocp0000056

Benavides, F. (2021). Causalidad y responsabilidad en salud laboral. Gaceta Sanitaria, 35(5), 502-505. https://dx.doi.org/10.1016/j.gaceta.2020.03.005

Boivin, D. B., & Boudreau, P. (2021). Disturbance of the circadian system in shift work and its health impact. *Nature and Science of Sleep*, *13*, 1-19. https://doi.org/10.2147/NSS.S268710

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301–328). Springer. https://doi.org/10.1007/978-94-007-6094-3_15

Bongers, P. M., de Winter, C. R., Kompier, M. A., & Hildebrandt, V. H. (2006). Psychosocial factors at work as risk factors for musculoskeletal disorders: A systematic review. *Scandinavian Journal of Work, Environment & Health, 32*(6), 431–443. https://doi.org/10.5271/sjweh.1040

Bovenzi, M. (2015). Exposure-response relationship for hand-arm vibration and sensorineural hearing loss. *International Archives of Occupational and Environmental Health, 88*(5), 587–595. https://doi.org/10.1007/s00420-014-0982-0

Breslow, N. E., & Day, N. E. (1980). *Statistical methods in cancer research. Volume 1—The analysis of case-control studies*. IARC Scientific Publications No. 32.

International Agency for Research

Checkoway, H., Pearce, N., & Kriebel, D. (2004). *Research methods in occupational epidemiology* (2nd ed.). Oxford University Press.

Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., & Poole, C. (2010). Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology, 39*(2), 417–420. https://doi.org/10.1093/ije/dyp334

Da Costa, B. R., & Vieira, E. R. (2010). Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. *American Journal of Industrial Medicine, 53*(2), 181–191. https://doi.org/10.1002/ajim.20760

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Digitale, J. C., Martin, J. N., Glymour, M. M., & Riley, L. W. (2022). Tutorial on directed acyclic graphs. *Journal of Clinical Epidemiology, 142,* 264-267. https://doi.org/10.1016/j.jclinepi.2021.08.001

Evans, D., Chaix, B., Lobbedez, T., Verger, C., & Flahault, A. (2012). Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. *BMC Medical Research Methodology, 12,* 156. https://doi.org/10.1186/1471-2288-12-156

Fox, M., Arah, O., Swanson, S., y Viallon, V. (2024). Causal diagrams to evaluate sources of bias. In De González, A., Richardson, D., y Schubauer-Be-

rigan M., (editors). Statistical Methods in Cancer Research Volume V: Bias Assessment in Case–Control and Cohort Studies for Hazard Identification. International Agency for Research on Cancer. Lyon (FR). https://www.ncbi.nlm.nih.gov/books/NBK612863/

Gifkins, J., Johnston, A., & Loudoun, R. (2020). Fatigue and recovery in shiftworking nurses: A scoping literature review. *International Journal of Nursing Studies, 111*, 103643. https://doi.org/10.1016/j.ijnurstu.2020.103643

Glymour, M. M., Weuve, J., & Chen, J. T. (2019). Methodological challenges in causal inference for social epidemiology. *Annual Review of Public Health, 40*, 221–238. https://doi.org/10.1146/annurev-publhealth-040218-044131

Gomstyn, A., y Jonker, A. (2024). What is a directed acyclic graph (DAG)? IBM.com. https://www.ibm.com/think/topics/directed-acyclic-graph

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology, 10*(1), 37–48. https://doi.org/10.1097/00001648-199901000-00008

Greenland, S., Pearl, J., & Robins, J. M. (2016). Causal diagrams for epidemiologic research: 15 years later. *Epidemiology, 27*(5), 615–617. https://doi.org/10.1097/EDE.0000000000000508

Gutiérrez -Strauss, A., y el Ministerio de la Protección Social de Colombia (2011). Guía técnica para el análisis de exposición a factores de riesgo ocupacional en el proceso de evaluación para la calificación de origen de la enfermedad profesional. Imprenta Nacional de Colombia. https://comunicandosalud.com/wp-content/uploads/2019/06/guia_exposicion_factores_riesgo_ocupacional.pdf

Hernán, M., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology, 15*(5), 615–625. https://doi.org/10.1097/01.ede.0000135174.63482.43

Hernán, M., Sterne, J., Higgins, J., Shrier, I., y Hernández-Diaz, S. (2024). A structural description of biases that generate immortal time. Epidemiology, Article (10). https://doi.org/10.1097/EDE.0000000000001808

Hernán, M., y Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

Hernández-Sampieri, R. y Mendoza, C. (2018). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta. McGraw-Hill Interamericana. https://bellasartes.upn.edu.co/wp-content/uploads/2024/11/METODOLOGIA-DE-LA-INVESTIGACION-Sampieri-Mendoza-2018.pdf

Hoxie, C. E., Parks, C. G., & Sandler, D. P. (2019). Occupational exposures and chronic obstructive pulmonary disease: A systematic review. *Occupational and Environmental Medicine, 76*(5), 334–342. https://doi.org/10.1136/oemed-2018-105367

Instituto Nacional de Seguridad e Higiene en el Trabajode España [INSST] (2026). Conceptos generales de la prevención de riesgos laborales y ámbito jurídico. https://www.insst.es/documents/94886/4154780/Tema%2022.%20Epidemiolog%C3%ADa%20laboral.pdf

International Labour Organization. (2021). *Guidelines on occupational safety and health management systems* (2nd ed.). ILO. https://www.ilo.org/safework/info/publications/WCMS_652225/lang--en/index.htm

Kalkanis, A., Voulgaris, S., & Steiropoulos, P. (2023). Recovery from shift work: Challenges and solutions. *Sleep Medicine Clinics*, *18*(3), 397-409. https://doi.org/10.1016/j.jsmc.2023.05.008

Keil, A. P., Cole, S. R., & Edwards, J. K. (2018). A practical guide to g-methods for time-varying exposures. *American Journal of Epidemiology, 187*(11), 2473–2481. https://doi.org/10.1093/aje/kwy163

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). Guilford Publications.

Krieger, N. (2011). *Epidemiology and the people's health: Theory and context.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195373066.001.0001

Malone, H., Hejazi, N. S., Blume, J. D., & van der Laan, M. J. (2021). ggDAG: Analyze and plot directed acyclic graphs using R and ggplot2. *Journal of Open Source Software, 6*(64), 3380. https://doi.org/10.21105/joss.03380

Mucci, T. (2 de diciembre de 2025). ¿En qué consiste la toma de decisiones basada en datos? IBM.com. https://www.ibm.com/es-es/think/topics/data-driven-decision-making

Nieuwenhuijsen, M. J. (2003). Exposure assessment in occupational and environmental epidemiology. *Occupational and Environmental Medicine, 60*(5), 313–314. https://doi.org/10.1136/oem.60.5.313

Organización Internacional del Trabajo [OIT] (2010). Lista de enfermedades profesionales de la OIT (revisada en 2010). https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40ed_protect/%40protrav/%40safework/documents/publication/wcms_125164.pdf

Organización Mundial de la Salud [OMS] (7 de noviembre de 2022). Occupational health: health workers. https://www.who.int/news-room/fact-sheets/detail/occupational-health--health-workers

Palmer, K. T., Harris, E. C., & Coggon, D. (2019). Carpal tunnel syndrome and its relation to occupation: A systematic literature review. *Occupational Medicine, 69*(1), 1–10. https://doi.org/10.1093/occmed/kqy145

Parks, C. G., D'Aloisio, A. A., DeRoo, L. A., & Sandler, D. P. (2019). Occupational exposures and chronic obstructive pulmonary disease in women. *American Journal of Respiratory and Critical Care Medicine, 199*(5), 612–620. https://doi.org/10.1164/rccm.201804-0732OC

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge University Press.

Pérez de Arenaza, D., y Pizarro, R. (2021). Sesgo de inmortalidad en estudios observacionales. Acta Gastroenterológica Latinoamericana, 51 (2),125-127. DOI: https://doi.org/10.52787/ajuj5684

Perkovic, E., Textor, J., Kalisch, M., & Maathuis, M. H. (2015). A complete generalized adjustment criterion. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence (UAI 2015)* (pp. 682-691). AUAI Press.

Poole, C. (2010). On the origin of causal inference. *Epidemiology, 21*(2), 133–135.

https://doi.org/10.1097/EDE.0b013e3181c9f4f7

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Rodríguez-García, M. Y Hoehndorf, R. (2018). Inferring ontology graph structures using OWL reasoning. BMC Bioinformatics 19,(7) . https://doi.org/10.1186/s12859-017-1999-8

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning, 45*(5-6), 341–358. https://doi.org/10.1016/j.lrp.2012.09.010

Rodríguez-Villamizar, L. (2017). Inferencia causal en epidemiología. Revista de Salud Pública, 19(3), 409-415. https://doi.org/10.15446/rsap.v19n3.66180

Robins, J. M. (1987). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling, 7*(9-12), 1393–1512. https://doi.org/10.1016/0270-0255(87)90001-0

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Sánchez-Niubò, A., Forero, C., y Benavides, F. (2016). Utilidad de los diagramas causales para conceptualizar mecanismos en epidemiología laboral. Archivos de Prevención de Riesgos Laborales, 19(2), 103-106. https://

dx.doi.org/10.12961/aprl.2016.19.02.4

Schaufeli, W. B., & Taris, T. W. (2014). A critical review of the Job Demands-Resources Model: Implications for improving work and health. *En G. F. Bauer & O. Hämmig (Eds.), Bridging occupational, organizational and public health: A transdisciplinary approach* (pp. 43–68). Springer. https://doi.org/10.1007/978-94-007-5640-3_4

Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology, 20*(4), 488–495. https://doi.org/10.1097/EDE.0b013e3181a819a1

Schoenbach, V. (2004). Inferencia causal. Revista Epidemiolog.net. 269-287.

Shrier, I., & Platt, R. W. (2008). Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology, 8*, 70. https://doi.org/10.1186/1471-2288-8-70

Smucler, E., Sapienza, F., & Rotnitzky, A. (2022). Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika, 109*(1), 49-65.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). MIT Press.

Steenland, K., & Ward, E. (2014). Asbestos: A review of the literature and evaluation of the evidence. *Journal of Occupational and Environmental Medicine, 56*(10), 1041–1048. https://doi.org/10.1097/JOM.0000000000000276

Textor, J., van der Zander, B., Gilthorpe, M. S., Liskiewicz, M., & Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology, 45*(6), 1887–1894. https://doi.org/10.1093/ije/dyw341

Universidad Catolica San Pablo [UCSP] (2026). Conoce qué es un factor de riesgo en salud ocupacional. https://postgrado.ucsp.edu.pe/articulos/factor-riesgo-salud-ocupacional/

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199325870.001.0001

VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology, 34*(3), 211–219. https://doi.org/10.1007/s10654-019-00494-6

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine, 167*(4), 268–274. https://doi.org/10.7326/M16-2607

VanderWeele, T. J., & Hernán, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference, 1*(1), 1–20. https://doi.org/10.1515/jci-2012-0002

VanderWeele, T. J., & Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics, 67*(4), 1406–1413. https://doi.org/10.1111/j.1541-0420.2011.01619.x

VanderWeele, T. J., Mathur, M. B., & Ding, P. (2020). Causal inference and data fusion in meta-analysis. *Annual Review of Statistics and Its Application, 7*, 345–368. https://doi.org/10.1146/annurev-statistics-031219-041034

Velarde-Martíne, A. (2018). Generación de Grafos Dirigidos Acíclicos mediante Métodos Aleatorios para la Planificación de Tareas. Conciencia Tecnológica, 55. 2018. https://www.redalyc.org/journal/944/94455712004/

Velasco-Cobos, L. (2024). El Tiempo como Entidad Ontológica: Una Exploración Filosófica de su Naturaleza y Validez en la Filosofía de la Ciencia[-Trabajo de Grado, Universidad Nacional Abierta y a Distancia, UNAD, Bogota-Colombia]. https://repository.unad.edu.co/bitstream/handle/10596/66916/lavelascoc.pdf?sequence=3&isAllowed=y

Vyas, M. V., Garg, A. X., Iansavichus, A. V., Costella, J., Donner, A., Laugsand, L. E., Janszky, I., Mrkobrada, M., Parraga, G., & Hackam, D. G. (2012). Shift work and vascular events: Systematic review and meta-analysis. *BMJ, 345*, e4800. https://doi.org/10.1136/bmj.e4800

Wang, X. S., Armstrong, M. E., Cairns, B. J., Key, T. J., & Travis, R. C. (2020). Shift work and chronic disease: The epidemiological evidence. *Occupational Medicine, 70*(4), 270–277. https://doi.org/10.1093/occmed/kqaa056

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

World Health Organization. (2021). *Occupational health: Guidelines for surveillance and risk assessment.* WHO Press. https://www.who.int/publications/i/item/9789240027029

Yang, G., Burgess, S. y Schooling, C. (2025).Illustrating the structures of bias from immortal time using directed acyclic graphs. International Journal of Epidemiology, 54(1). https://doi.org/10.1093/ije/dyae176


NOVO MUNDO