

Aplicación de grafos acíclicos dirigidos en la evaluación de un set mínimo de ajuste de confusores: un complemento al modelamiento estadístico en estudios epidemiológicos observacionales

FABIOLA WERLINGER^{1,2,a}, DANTE D. CÁCERES^{3,4,b}

Directed acyclic graphs in statistical modelling of epidemiological studies

Background: Confusion in observational epidemiological studies distorts the relationship between exposure and event. “Step by step” regression models, diverts the decision to a statistical algorithm with little causal basis. Directed Acyclic Graphs (DAGs), qualitatively and visually assess the confusion. They can complement the decision on confounder control during statistical modeling. **Aim:** To evaluate the minimum set of confounders to be controlled in a cause-effect relationship with the use of “step-by-step regression” and DAGs, in a study of arsenic exposure. **Material and Methods:** We worked with data from Cáceres et al., 2010 in 66 individuals from northern Chile. The interindividual variability in the urinary excretion of dimethyl arsenic acid attributable to the GSTT1 polymorphism was estimated. A causal DAG was constructed using DAGitty v2.3 with the list of variables. A multiple linear regression model with the step-by-step backwards methodology was carried out. **Results:** The causal diagram included 12 non-causal open pathways. The minimum adjustment set corresponded to the variables “sex”, “body mass index” and “fish and sea-food ingest”. Confusion retention of the multivariate model included normal and overweight status, gender and the interaction between “water intake” and GSTT1. **Conclusions:** The use of DAG prior to the modeling would allow a more comprehensive, coherent and biologically plausible analysis of causal relationships in public health.

(Rev Med Chile 2018; 146: 907-913)

Key words: Confounding Factors (Epidemiology); Epidemiologic Methods; Regression Analysis.

La epidemiología plantea que el fenómeno de enfermedad no se distribuye al azar¹ buscando determinar los factores o “causas” que condicionarían su aparición en la población²⁻⁴. Tal aproximación se ha complejizado al incorporarse conceptos como la multicausalidad⁵, el efecto en términos probabilísticos² y la inclusión de posibles confusores⁶. Estos últimos, también llamados

“factores de confusión” o “variables extrañas”, corresponden a cualquier condición que, teniendo una atribución causal², distorsionan la relación evaluada entre una exposición y su resultado, ya sea exagerando (confusión positiva), debilitando (confusión negativa) o, incluso, cambiando la dirección (confusión cualitativa) de la medida de efecto. Como requisito fundamental, estas

¹Centro de Epidemiología y Vigilancia de las Enfermedades Orales (CEVEO), Facultad de Odontología, Universidad de Chile. Santiago, Chile.

²Departamento de Tecnología Médica. Facultad de Medicina, Universidad de Chile. Santiago, Chile.

³Programa de Salud Ambiental, Escuela de Salud Pública, Universidad de Chile. Santiago, Chile.

⁴Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica, Chile.

^aMédico Veterinario, Magíster en Salud Pública, Magíster en Bioestadística©.

^bMédico Veterinario. Magíster en Salud Pública, Doctor en Medicina Preventiva y Salud Pública.

Fuente de apoyo financiero: sin apoyo financiero.

Los autores declaran no tener conflictos de interés.

Recibido el 5 de diciembre de 2017, aceptado el 6 de julio de 2018.

Correspondencia a:
Fabiola Werlinger Cruces
fwerlinger@med.uchile.cl

condiciones deben encontrarse asociadas tanto a la exposición como al evento, sin corresponder a un paso causal intermedio entre ellas⁶⁻⁸.

Para su control, cuando se consideran diversas variables, se proponen métodos multivariados⁸⁻¹² como la regresión “paso a paso hacia atrás” o “*backward selection*”, que aseguraría la selección de todos los confusores relevantes en la construcción del modelo¹³, restando secuencialmente variables a un modelo completo inicial en busca de uno más simple^{8,14-16}. Sin embargo, diversos autores proponen que este procedimiento excluiría confusores importantes o dejaría en el modelo “optimizado” variables erróneamente clasificadas como confusores, dado que la selección se realiza a través de algoritmos matemático-estadísticos basados en el valor p , donde la influencia del investigador es mínima^{6,7,14,17,18}.

Esto ha generado en la actualidad la propuesta de otros métodos alternativos más simples e ilustrativos como los grafos acíclicos dirigidos o DAGs (*Directed Acyclic Graphs*)¹⁰ que, a partir de formas gráficas construidas sobre el conocimiento experto *a priori* del investigador^{7,9}, facilitarían la comprensión causal del fenómeno y el tipo de vinculación entre las variables involucradas, minimizando la introducción de sesgos durante el diseño del estudio y en el análisis de resultados^{6,14,19}.

Estos DAGs, también conocidos como diagramas causales o grafos²⁰, permiten desplegar esquemáticamente relaciones de causalidad^{2,6,9,21,22,23} de una manera rigurosa y sistemática^{23,24} a nivel individual o poblacional^{2,6,9,12,20}, permitiendo no solo identificar posibles factores de confusión, sino que también describir el alcance de un problema de investigación o, incluso, planificar la recopilación de datos¹².

Su estructura está conformada por el enlace de variables componentes de un modelo causal denominadas por una letra (nodo), donde la dirección causal es representada por flechas (camino)^{21,23} como efecto directo de una variable sobre otra^{2,6,9,12,21,22}. El hecho de que ninguna variable pueda causarse a sí misma, directa o indirectamente, es lo que indica que un DAG sea acíclico^{12,25}. Un ejemplo se presenta en la Figura 1. En ella, la notación $A \rightarrow E$ indica que A es causa directa de E. D y Z son causalmente independientes, lo que significa que un cambio en cualquiera de ellas no afectará el valor de la otra^{24,26}. La línea punteada de Z representa una variable no medida o especifica-

da²¹ que actúa como causa indirecta de E. En este caso, E es causalmente dependiente de Z, pero si A no se modifica, Z no causará cambios en E^{9,23}. B correspondería a un factor intermedio entre A y E, mientras que C corresponde a una variable que se relaciona con ambos, es decir, un ancestro común. El efecto total de A sobre E puede ser estimado a partir de las distribuciones observadas del resto de las variables del diagrama causal: A, B, C, D y Z^{24,26}.

Un camino corresponderá a cualquier ruta que conecte la exposición (A) al evento (E), independiente de la dirección de la flecha^{9,24,27}. Si partiendo de la exposición es congruente en su dirección hasta llegar al evento ($A \rightarrow B \rightarrow E$), será entendido como “camino causal dirigido”^{9,12}. Si la vía incluye alguna flecha apuntando en dirección opuesta ($A \leftarrow C \rightarrow E$), corresponderá a un “camino de puerta trasera”⁹ o vía no causal^{12,27}. En el caso de $A \rightarrow B \leftarrow D \rightarrow E$, nos encontramos frente a un “camino de choque” o colisionador donde dos puntas de flecha “chocan” o se encuentran en una variable (B) denominada “*collider*”^{9,12,27}, indicando que no es necesario ajustar, pues la vía naturalmente está bloqueada²⁷.

En el control de confusión se evalúa la existencia de vías no casuales o “caminos de puerta trasera” abiertos que puedan transmitir una asociación espuria que sesgue la estimación de la asociación^{12,27}. Se busca un ajuste por la causa común que “cierre” todos los posibles caminos no causales^{9,27} entre la exposición y el evento²⁷. Un camino de confusión en nuestro ejemplo corresponde a $A \leftarrow C \rightarrow E$, donde debiera ajustarse por C,

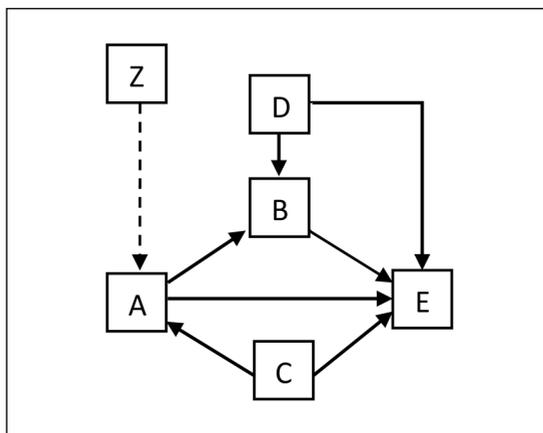


Figura 1. Estructura básica de un DAG. Fuente: Elaboración propia. A = exposición; E = evento; Z, B, C, D = covariables.

factor de confusión asociado con la exposición y el evento²⁴.

El objetivo de este trabajo fue contrastar, cualitativamente, dos métodos usados para la selección del set mínimo de confusores a controlar en una relación causal, con el uso de regresión “paso a paso hacia atrás” y DAGs, aplicados a un estudio epidemiológico observacional donde se evalúa la exposición a arsénico a través del agua (causa) sobre la excreción de metabolitos arsenicales en orina (efecto).

Material y Método

Con la autorización previa de todos sus autores, se trabajó con los datos del modelamiento realizado por Cáceres et al., 2010²⁸ en una muestra de 66 participantes crónicamente expuestos a arsénico en agua potable, residentes en la ciudad de Antofagasta, en el norte de Chile. Se estimó el efecto de la variabilidad interindividual atribuible a la presencia del alelo de la enzima glutatión S-transferasa T1 (GSTT1), sobre la excreción urinaria del ácido dimetilarsénico (ADMA), metabolito detoxificador de la exposición a arsénico por consumo de agua²⁹. Paralelamente, se realizó una encuesta de exposición ambiental, de características antropométricas y de estilo de vida.

Construcción del DAG

A partir de las variables levantadas en este estudio, se construyó un DAG causal inicial según lo sugerido por Textor J., con el *software* DAGitty

v2.3³⁰ (Figura 2). Se especificó la nomenclatura de cada variable y sus relaciones con el evento según la evidencia científica¹² (Tabla 1). En esta etapa, la variante polimórfica de la enzima de GSTT1 fue incorporada como variable modificadora del efecto directo de la exposición a arsénico (“ingesta de agua”) sobre la excreción urinaria del ADMA (resultado), de acuerdo a lo sugerido por Weinberg (2007)³¹.

La evaluación del DAG final se desarrolló con la metodología adaptada de Shrier and Platt's, 2008¹⁹, con un set mínimo de ajuste (SMA) sin pasos intermedios entre la exposición y el resultado, eliminando variables que no intervienen en la relación causal e identificando probables asociaciones inducidas por el ajuste de colisionadores^{19,27}.

Modelamiento estadístico

En una segunda etapa, utilizando el *software* Stata 11.0[®], se construyeron 2 modelos de regresión lineal múltiple: uno con el SMA resultante del DAG causal y otro con la metodología de selección “paso a paso hacia atrás” ($\alpha = 0,1$)⁸. En el caso de la “ingesta de agua” e “índice de masa corporal (IMC)” se construyeron variables *dummy* considerando como referencia la categoría de menor magnitud. Adicionalmente, se evaluó un tercer modelo con la relación basal del problema de investigación (sin confusores). En todos, la variable resultado correspondió al ADMA, la variable de exposición a la “ingesta de agua” y el resto a covariables del modelo, agregando en cada uno el efecto de interacción “GSTT1*ingesta de agua”^{18,27}.

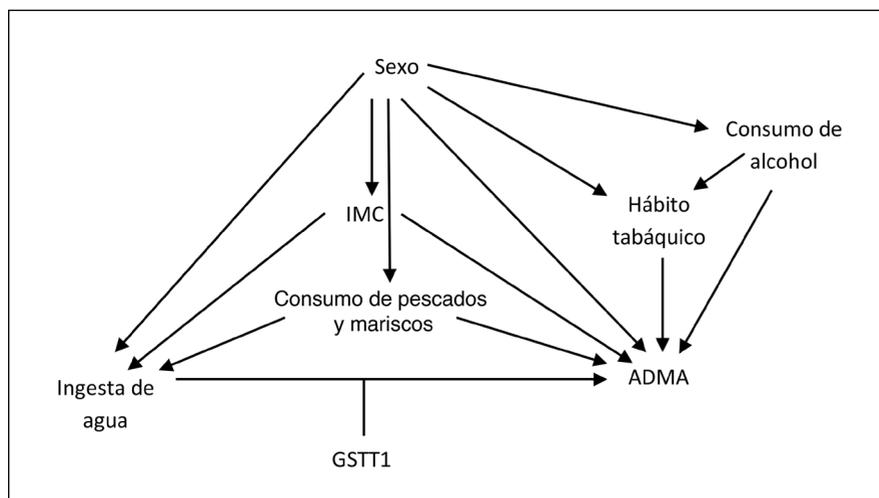


Figura 2. Grafo acíclico dirigido propuesto para el mecanismo causal entre la ingesta de agua y la excreción de ADMA modificado por GSTT1. Fuente: imagen adaptada de DAGitty v2.3. Elaboración propia.

Tabla 1. Variables incluidas en el problema de investigación

Variables	Escala de medición	Referencia
<i>Variable exposición</i> Ingesta de agua	No consume < 1 vaso/día 1-2 vasos/día > 2 vasos/día	Vasken et al., 2000; Cáceres et al., 2005; Macedo-Ojeda et al., 2008
<i>Variable resultado</i> Ácido dimetilarsénico (ADMA)	% en orina (0 a 100%)	Vasken et al., 2000; Cáceres et al., 2005
<i>Covariables</i> GSTT1 (modificador)	Presente /Ausente	Chiou et al., 1997; Kile et al., 2005; Liu et al., 2001
Sexo	Hombre / Mujer	Hopeynhayn-Rich et al., 1996; Yu Mei Hsueh, 2002; Lindberg et al., 2008
Índice de masa corporal	Bajo peso (< 18,5) Normo peso (18,5-24,9) Sobrepeso (\geq 25)	Kile et al., 2005; Medlineplus 2016
Hábito tabáquico	nº de cigarrillos fumados el día anterior (0 a 50)	Hopeynhayn-Rich et al., 1996; Romero et al., 2009; Conace 2005; Jané, 2004
Consumo de pescados y mariscos	sí/no (el día anterior)	Heinrich-Ramm et al., 2002; Benarroch & Pérez, 2011
Consumo de alcohol	sí/no (el día anterior)	Bao & Shi, 2010; Benarroch & Pérez, 2011; Conace, 2005; Galán et al., 2014; Mora-Ríos & Natera, 1999

GSTT1: Glutation S-Transferasa T1.

Resultados

En la Figura 2, se dispone el DAG final construido para la relación evaluada.

El diagrama incluyó 14 “camino de puerta trasera” o vías no causales, de los cuales 2 estaban bloqueados por un colisionador (hábito tabáquico):

1. Ingesta de agua \leftarrow **sexo** \rightarrow ADMA.
2. Ingesta de agua \leftarrow **IMC** \rightarrow ADMA.
3. Ingesta de agua \rightarrow **consumo de pescados y mariscos** \rightarrow ADMA.
4. Ingesta de agua \leftarrow **sexo** \rightarrow **IMC** \rightarrow ADMA.
5. Ingesta de agua \leftarrow **IMC** \leftarrow **sexo** \rightarrow ADMA.
6. Ingesta de agua \leftarrow **sexo** \rightarrow **consumo de pescados y mariscos** \rightarrow ADMA.
7. Ingesta de agua \leftarrow **sexo** \rightarrow consumo de alcohol \rightarrow ADMA.
8. Ingesta de agua \leftarrow **IMC** \leftarrow **sexo** \rightarrow consumo de alcohol \rightarrow ADMA.
9. Ingesta de agua \leftarrow **sexo** \rightarrow hábito tabáquico \rightarrow ADMA.
10. Ingesta de agua \leftarrow **IMC** \leftarrow **sexo** \rightarrow hábito tabáquico \rightarrow ADMA.

11. Ingesta de agua \leftarrow **IMC** \leftarrow **sexo** \rightarrow consumo de alcohol \rightarrow hábito tabáquico \rightarrow ADMA.

12. Ingesta de agua \leftarrow **sexo** \rightarrow consumo de alcohol \rightarrow hábito tabáquico \rightarrow ADMA.

13. Ingesta de agua \leftarrow **sexo** \rightarrow hábito tabáquico \leftarrow consumo de alcohol \rightarrow ADMA.

14. Ingesta de agua \leftarrow **IMC** \leftarrow **sexo** \rightarrow hábito tabáquico \leftarrow consumo de alcohol \rightarrow ADMA.

En los 12 caminos no casuales abiertos o “camino de puerta trasera” que introducen confusión en el DAG, comprendieron las variables “sexo”, “IMC” y “consumo de pescados y mariscos”, constituyendo el SMA (Figura 3). Por lo tanto, en el modelamiento estadístico de esta relación causal debiera incluir el ajuste por estos confusores.

Los confusores retenidos con la metodología “paso a paso hacia atrás” correspondieron al “sexo”, las categorías “normopeso” y “sobrepeso” y la interacción “GSTT1*ingesta de agua”. Los modelos construidos con ambas metodologías y sus confusores seleccionados se resentan en la Tabla 2.

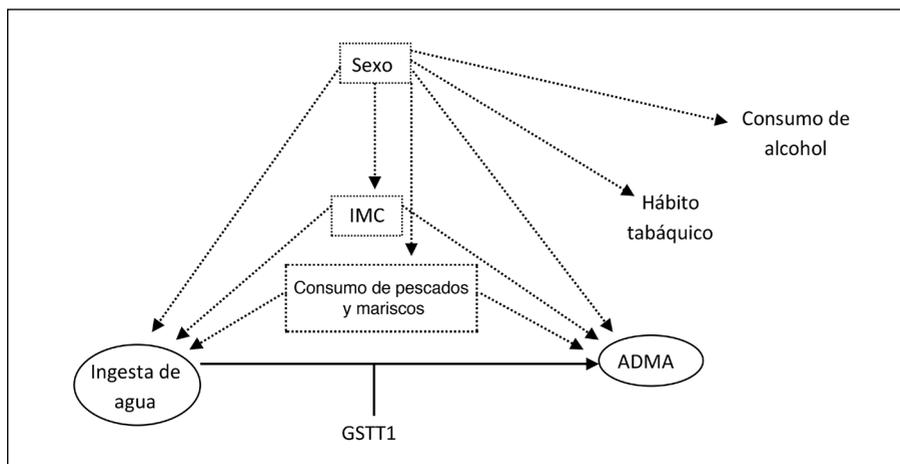


Figura 3. DAG con set mínimo de ajuste (indicado en cajas). Fuente: imagen adaptada de DAGgitty v2.3. Elaboración propia.

Tabla 2. Modelos finales ajustados con uso de Dag y selección “paso a paso hacia atrás” en la evaluación de la excreción de ADMA urinario y polimorfismo GSTT1

Modelo con SMA de DAG	Modelo de selección “paso a paso hacia atrás”
(1) GSTT1	(1) GSTT1*ingesta > 2 vasos/día
(2) GSTT1*ingesta > 2 vasos/día	(2) Sexo
(3) Ingesta < 1 vaso/día	(3) Normopeso
(4) Ingesta 1-2 vasos/día	(4) Sobre peso
(5) Ingesta > 2 vasos/día	
(6) Sexo	
(7) Normopeso	
(8) Sobre peso	
(9) Bajo peso	
(10) Consumo de pescados y mariscos	

GSTT1: Glutacion S-transferasa T1; SAM: Set de Ajuste Mínimo.

Discusión

Esta propuesta aplicó dos aproximaciones metodológicas utilizadas en el control del efecto de variables confusoras sobre una relación causa-efecto de un problema de salud pública: el análisis de regresión con selección “paso a paso hacia atrás” y el uso de DAGs. La discusión de nuestros resultados se propone desde la perspectiva de la selección de los confusores a ajustar en el modelamiento estadístico del fenómeno en estudio.

A la luz de los resultados, se observa cierta concordancia entre ambos procedimientos al se-

leccionar variables comunes como “sexo” e “índice de masa corporal”, que desde una perspectiva de la coherencia y la plausibilidad biológica tiene sentido. Ambos factores corresponden a causas comunes entre la exposición y el evento evaluado^{2,9}, siendo responsables de parte del porcentaje del ADMA urinario presente en los individuos en estudio. La diferencia sustancial radica en que si bien el modelo de selección “paso a paso hacia atrás” considera relevante la interacción entre el polimorfismo genético y la ingesta de agua (GSTT1*ingesta de agua), omite por naturaleza matemática (valor $p > 0,1$)^{8,15} el efecto aislado del

mismo polimorfismo, lo que aconteció igualmente cuando se evaluaron los modelos mencionados con diferentes puntos de corte del valor p (0,05 y 0,2), sin evidenciarse un cambio en las variables seleccionadas. No seleccionar esta variable estructural en el modelamiento^{18,32} impediría considerar la inclusión de cualquier interacción en la que ella participe, principio básico de un modelamiento jerárquico bien estructurado¹⁸. En nuestro caso, se traduce en que la presencia de la interacción entre la ingesta del agua y el polimorfismo genético en el modelo de selección no tendría plausibilidad biológica. La exclusión de parte de esta condición refleja su imposibilidad de discriminar el comportamiento de los eventos o condiciones biológicas en los organismos vivos. Si consideramos que nuestro DAG representa de manera efectiva la totalidad de las relaciones causales y no casuales para el efecto de interés, observamos que nos permite una mirada más comprehensiva de los modelos de regresión aplicados de manera combinada previo al proceso de ajuste. En nuestro caso resguarda que, independiente del tamaño muestral, la participación del “sexo”, el “índice de masa corporal” y el “consumo de pescados y mariscos” en los individuos, necesariamente se acompañe de las variables estructurales de la relación¹⁸ y las posibles interacciones en las que ellas participan³², armonizando de manera más respetuosa con la concepción de causalidad del fenómeno^{9,11}.

De acuerdo con nuestros resultados es posible afirmar que las ventajas del uso de DAGs son interesantes como alternativa combinada al análisis estadístico de selección tradicional para el control y detección de confusión^{18,20}, permitiendo una observación más clara y económica de la vinculación entre las variables y sus relaciones, diferenciando las causales de las no causales^{9,12,23,27} e identificando mejor los factores de confusión¹² que no fueron capturados en el modelamiento convencional²¹. Esto toma particular relevancia en la medición en salud pública, que deriva sus conclusiones principalmente de estudios observacionales con un control limitado de las variables en los grupos de interés^{7,25}, posicionándolo como una herramienta de fácil manejo que apoye la planificación en la recolección, análisis^{9,23,27} y comunicación de resultados²¹.

Sin duda que deben resguardarse diversas consideraciones en la aplicación de esta metodología,

como la imposibilidad de establecer diferentes niveles de exposición, o que no considera la existencia de relaciones no lineales entre nuestras variables^{12,27,31}. Tampoco es posible estimar su distribución o transmitir la magnitud del riesgo por lo que la recomendación es combinarlos con los enfoques deductivos tradicionales¹² como en nuestro caso. También, aunque su construcción parece sencilla, el apoyo de *softwares* puede ser recomendable, teniendo en cuenta que no es posible plasmar de manera directa efectos de interacción que se presentan independientemente de la escala seleccionada como ocurre en vías metabólicas de desintoxicación específica³¹.

Agradecimientos: Profesores Mauricio Fuentes Albuquerque y Sergio Alvarado Orellana. Programa de Bioestadística, Escuela de Salud Pública. Facultad de Medicina, Universidad de Chile.

Referencias

1. López S, Hernández M. Desarrollo histórico de la epidemiología. En: Hernández M, editor. *Epidemiología diseño y análisis de estudio*. 2ª ed. México. Panamericana. 2009.
2. Pearl J. An introduction to causal inference. International J Biostat [internet] 2010 [acceso: enero de 2015]; 6 (2): Article 7. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836213/pdf/ijb1203.pdf>
3. Rothman KJ. Causes. Am J Epidemiol 1976; Vol 104 (6): 587-92.
4. Parascandola M. Causes risk and probabilities: Probabilistic concepts of causation in chronic disease epidemiology. Prev Med 2011; 53: 232-4.
5. Mc Mahon B, Pugh T. *Epidemiology; principles and methods*. Boston: 2ª ed. 1970.
6. Rothmann K, Greenland S, Poole C, Lash T. *Modern epidemiology*. 3ª ed. Lippincott Williams & Wilkins. Philadelphia, EE.UU. 2008.
7. Szklo M, Nieto J. *Epidemiology: beyond the basics*. 3ª ed. Jones & Barlett Learning. Burlington, Mass. EE.UU. 2012.
8. Dupont W. *Statistical modeling for biomedical researchers*. 2ª edición. Ed. Cambridge University Press. Cambridge, Reino Unido. 2009.
9. Hernán M, Hernández-Díaz S, Werler M, Mitchell A. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol 2002; 155 (2): 176-84.

10. Kestenbaum B. Methods to control for confounding. En: Adeney K, Weiss N, editores. *Epidemiology and Biostatistics. An Introduction to Clinical Research*. 1ª ed. Springer; Nueva York, EE.UU. 2009.
11. Suttorp M, Siegerink B, Jager K, Zoccali C, Dekker F. Graphical presentation of confounding in directed acyclic graphs. *Nephrol Dial Transplant* 2015; 30 (9): 1418-23.
12. Brewer L, Wright J, Rice G, Neas L, Teuschler L. Causal inference in cumulative risk assessment: The roles of directed acyclic graphs. *Environ Int* 2017; 102: 30-41.
13. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986; 15 (3): 413-9.
14. Dunkler D, Plischke M, Leffondré K, Heinze G. Augmented backward elimination: pragmatic and purposeful way to develop statistical models. *Plos One* [internet] 2014 [acceso: junio de 2015]; 9 (11): e113677. Disponible en: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0113677>.
15. Reimann C, Filzmoser P, Garret R, Dutter R. *Statistical data analysis explained: applied environmental statistics with R*. Vienna University of Technology. John Wiley & Sons. 2008.
16. Silva L. Selección algorítmica en las aplicaciones biomédicas de la regresión múltiple. *Med Clin* 2001; 116: 741-45.
17. Thiese M, Arnold Z, Walker S. The misuse and abuse of statistics in biomedical research. *Biochemia Medica* 2015; 25 (1): 5-11.
18. Kleinbaum D, Klein M. Modeling Strategy Guidelines. En: Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editores. *Logistic Regression. A Self-Learning Text*. 3ª edición. Springer. Nueva York, EE.UU. 2010. P. 162-89.
19. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* [revista en internet]; 2008 [acceso en octubre de 2016]; 8:70. Disponible en: <http://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-8-70>.
20. Tu YK, Gilthorpe M. Path. capítulo: Directed Acyclic Graphs. En: *Statistical Thinking in Epidemiology*. 1ª ed. CRC Press. Nueva York, EE.UU. 2012.
21. Greenland S, Pearl J, Robins J. Causal diagrams for epidemiologic research. *Epidemiology* 1999; 10(1):38-48.
22. Shahar E, Shahar D. Causal diagrams and the cross-sectional study. *Clinical Epidemiology* [internet] 2013 [acceso: marzo de 2015]; 5: 57-65. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3600935/>
23. Glaymor M. Using causal diagrams to understand common problems in social epidemiology. En: Oakes M Kaufman J editores. *Methods in Social Epidemiology*. Wiley & Sons. Hoboken, NJ, EE.UU. 2006. p: 387-422.
24. Shepley B. *Cause and correlation in biology*. University Press. Cambridge, Reino Unido. 2009.
25. Hernan M, Hernández-Díaz, S Robin J. A structural approach to selection bias *Epidemiology* 2004; 15 (5): 615-25.
26. Pearl J. Causality. Introduction to probabilities graphs and causal models. [monografía en Internet] 1999 [acceso: julio de 2015]. Disponible en: <http://bayes.cs.ucla.edu/BOOK-99/book-toc.html>
27. Williamson J, Aitken Z, Lawrie J, Dharmage S, Burgess J, Forbes A. Introduction to causal diagrams for confounder selection. *Respirology* 2014; 19: 303-11.
28. Cáceres D, Werlinger F, Orellana M, Jara M, Rocha R, Alvarado S, et al. Polimorfism of Glutathione-S-Transferase (GST) variants and its effects on distribution of urinary arsenic species in people exposed to low inorganic arsenic in tap water: an exploratory study. *Arch Environ Occup Health* 2010; 65 (3): 140-7.
29. Chiou HY, Hsueh YM, Hsie LL, et al. Arsenic methylation capacity, body retention, and null genotypes of glutathione S-transferase M1 and T1 among current arsenic-exposed residents in Taiwan. *Mutat Res* 1997; 386 (3): 197-207.
30. Textor J. Drawing and analyzing causal DAGs with DAGitty user manual for version 2.2. [monografía en internet]. Utrecht: Johannes Textor. 2014. [acceso en junio de 2015] Disponible en: <http://www.dagitty.net/>
31. Winberg C. Can DAGs Clarify Effect Modification?. *Epidemiology* [revista en internet] 2007 [acceso julio de 2016]; 18 (5): 569-72. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2235194/>
32. Jung S. Challenges in epidemiological and statistical evaluations of effect modifiers and confounders. *Epidemiology* [revista en internet]; 2014 [acceso en octubre de 2016]; 2 (277): 1-2. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4261598/>