



Universidade do Estado do Rio de Janeiro

Centro Biomédico

Instituto de Medicina Social

Táisa Rodrigues Cortes

**Utilização de diagramas causais e a comunicação de fontes de
incerteza em estudos observacionais**

Rio de Janeiro

2018

Táisa Rodrigues Cortes

Utilização de diagramas causais e a comunicação de fontes de incerteza em estudos observacionais

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Saúde Coletiva, da Universidade do Estado do Rio de Janeiro. Área de concentração: Epidemiologia.

Orientador: Prof. Dr. Claudio José Struchiner

Rio de Janeiro

2018

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/CB/C

C828 Cortes, Taísa Rodrigues
Utilização de diagramas causais e a comunicação de fontes de incerteza
em estudos observacionais / Taísa Rodrigues Cortes – 2019.
174 f.

Orientador: Claudio José Struchiner

Tese (doutorado) – Universidade do Estado do Rio de Janeiro, Instituto
de Medicina Social.

1. Estudos observacionais como assunto – Teses. 2. Causalidade–
Teses. 3. Características dos estudos epidemiológicos – Teses. 4. Incerteza
– Teses. 5. Análise estatística – Teses. 6. Estatística como assunto – Teses.
7. Epidemiologia – Teses. 8. Literatura de revisão como assunto – Teses.
I. Struchiner, Claudio José. II. Universidade do Estado do Rio de Janeiro.
Instituto de Medicina Social. III. Título.

CDU 616-036.22:311

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta
dissertação, desde que citada a fonte.

Assinatura

Data

Táisa Rodrigues Cortes

Utilização de diagramas causais e a comunicação de fontes de incerteza em estudos observacionais.

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Saúde Coletiva, da Universidade do Estado do Rio de Janeiro. Área de concentração: Epidemiologia.

Aprovada em 29 de maio de 2018.

Banca Examinadora: _____

Prof. Dr. Claudio José Struchiner (Orientador)

Instituto de Medicina Social – UERJ

Prof. Dr. Guilherme Loureiro Werneck

Instituto de Medicina Social – UERJ

Prof. Dr. Michael Eduardo Reichenheim

Instituto de Medicina Social – UERJ

Prof. Dr. José Ueleres Braga

Instituto de Medicina Social – UERJ

Prof. Dr. Antônio José Leal Costa

Universidade Federal do Rio de Janeiro

Rio de Janeiro

2018

AGRADECIMENTOS

Gostaria de expressar a minha admiração e gratidão ao meu orientador pelo aprendizado constante, encorajamento e amizade.

Agradeço imensamente ao meu pai por todo amor, apoio e amizade.

À toda comissão da banca pelas sugestões proveitosas.

Aos professores e professoras do Instituto de Medicina Social, em especial ao Michael Reichenheim, Guilherme Werneck e Washington Junger.

Aos meus amigos da UERJ, em especial a Karine Boclin, Ana Paula Razal, Andre de Albuquerque e André Moreira.

A todos os funcionários do Instituto de Medicina Social.

RESUMO

CORTES, Taísa Rodrigues. **Utilização de diagramas causais e a comunicação de fontes de incerteza em estudos observacionais**. 2018. 174 f. Tese (Doutorado em Saúde Coletiva) – Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

Os diagramas causais (gráficos acíclicos direcionados – DAG) têm sido apontados como uma das principais ferramentas que podem contribuir para a qualidade metodológica e do relato de estudos observacionais. No entanto, pouco se sabe sobre como essas ferramentas têm sido utilizadas nas investigações empíricas. Neste trabalho, foi realizada uma revisão da literatura com o objetivo de descrever o quanto, como e onde os diagramas causais têm sido utilizados em estudos observacionais analíticos nos últimos 18 anos. Foram realizadas buscas por citações e pesquisas por palavras-chaves nas bases de dados do PubMed e Web of Science. Em uma amostra de 100 artigos que apresentaram a estrutura causal, foram avaliadas as características dos modelos gráficos e o relato de confundimento. Além disso, foi realizada uma análise comparativa do relato das limitações do estudo e da frequência de marcadores linguísticos de incerteza (*hedges*) nos artigos com e sem a utilização de DAG causais. Foram identificadas 1034 publicações, totalizando 5021 autores e 85 países de afiliação. Apenas 430 artigos (42%) forneceram a estrutura gráfica. A maioria das publicações contém apenas um DAG causal (87%) e poucos modelos gráficos contêm confundidores não observados (23%), ou a representação de erros de mensuração (6%) e mecanismos de seleção (3%). O relato de modificações no conjunto de ajuste foi observado em 19% das publicações. Além disso, 20% foram classificadas como possível ocorrência da falácia da tabela 2. O número de limitações do estudo reconhecidas pelos autores e a frequência de marcadores de incerteza foram semelhantes nas amostras de artigos com e sem diagramas causais. No entanto, o relato de avaliações quantitativas das limitações do estudo foi mais frequente entre os artigos com DAG (52% vs. 21%). Há necessidade de mais discussões e estudos sobre a construção e análise de modelos causais e o desenvolvimento de recomendações gerais para apresentação de DAG causais nos artigos científicos.

Palavras-chave: Diagramas causais. Gráficos acíclicos direcionados. Relatórios de pesquisa. Confundimento. Limitações do estudo. Hedges. Incerteza.

ABSTRACT

CORTES, Taísa Rodrigues. **Use of causal DAG and reporting of uncertainty in observational studies.** 2018. 174 f. Tese (Doutorado em Saúde Coletiva) – Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2018.

Causal diagrams (Directed Acyclic Graphs- DAG) have been suggested as a potential tool to improve methodological and reporting quality of observational studies. However, little is known about their empirical applications. A literature review was performed to describe when, where and how causal diagrams have been applied in analytical observational studies in the past 18 years. Potential studies were retrieved through a citation search and a keyword search within PubMed and Web of Science. For a sample of 100 articles that provided the causal structure, the graphical model's features, and reporting of confounding were assessed. In addition, reporting limitations and use of uncertainty markers (hedges) were compared between articles with and without causal DAG. From 2000 to 2017, 1034 publications were identified, including 5021 authors and 85 country affiliations. Only 430 articles (42%) provided the graphic structure. The majority of studies contain only one DAG, and few graphical models included unobserved confounders (23%), or measurement errors (6%) and selection mechanisms (3%). Modifications in the adjustment set were reported in 19% of the articles. Besides, 20% were classified as a possible occurrence of Table 2 fallacy. The number of study limitations acknowledged by the authors and the frequency of uncertainty markers was similar in articles with and without causal diagrams. However, the reporting of quantitative assessments of study limitations were more frequent among studies with DAG (52% vs. 21%). There is a need for further discussion and research concerning the construction and analysis of causal models and the development of reporting guidelines of causal DAG in empirical studies.

Keywords: Causal diagram. Directed acyclic graph. Reporting. Confounding. Study limitations. Hedges. Uncertainty.

LISTA DE FIGURAS

Figura 1 - Exemplo de uma intervenção $do(x)$ em um DAG causal.....	24
Figura 2 - Representação estrutural de confundimento.	27
Figura 3 - Estratégias de controle de confundimento.	27
Figura 4 - Representação estrutural de erros de mensuração.	29
Figura 5 - Representação estrutural de erros de mensuração.	30
Figura 6 - Representação estrutural de erros de mensuração.	31
Figura 7 - Exemplo de viés de colisão (paradoxo de Berkson).	33
Figura 8 - Taxonomia de ignorância de Smithson.	42
Figura 9 - Fluxograma da seleção dos estudos.....	60
Figura 10 - Mapeamento de palavras-chave utilizadas nos títulos e resumos de artigos com DAG causal.....	70
Figura 11 - Diferença entre o número de variáveis de confusão incluídas no modelo final e o número de vértices do DAG.....	73
Figura 13 - Diferentes DAG e seus respectivos conjuntos de ajuste e a classe de modelos Markov equivalentes.	80

LISTA DE GRÁFICOS

Gráfico 1 - Número de publicações por ano de estudos observacionais analíticos com o relato da utilização de diagramas causais.	67
Gráfico 2 - Principais fontes de publicação dos artigos com diagramas causais.	69
Gráfico 3 - Tipo de limitações do estudo e fontes de viés reconhecidas pelos autores em artigos com e sem diagramas causais.....	92
Gráfico 4 - Diagrama causal representando as hipóteses de sobre a relação entre a apresentação de diagramas causais em estudos observacionais e o relato das limitações do estudo.	96
Gráfico 5 - Gráfico acíclico direcionado simplificado representando as hipóteses de seleção.	98
Gráfico 6 - Gráfico acíclico direcionado simplificado representando hipóteses alternativas de seleção.	99
Gráfico 7 - Gráfico acíclico direcionado simplificado representado as hipóteses alternativas de seleção e confundimento.	100
Gráfico 8 - Diagrama causal representado as hipóteses sobre a relação entre a apresentação de diagramas causais em estudos observacionais e o nível de atenuação da linguagem.	101

LISTA DE TABELAS

Tabela 1 - Hierarquia de perguntas	21
Tabela 2 - Exemplos de hedges utilizados em artigos científicos.	48
Tabela 3 - Principais países de afiliação dos coautores de publicações com o relato da utilização de diagramas causais.....	68
Tabela 4 - Aplicações e características dos diagramas causais (N=100).	72
Tabela 5 - Frequência de itens importantes na comunicação de confundimento em uma amostra de artigos com diagramas causais (N=100).	74
Tabela 6 - Características gerais dos artigos e dos coautores.....	90
Tabela 7 - Frequência de termos hedges e reconhecimento das limitações do estudo em artigos com e sem diagramas causais.	91

LISTA DE ABREVIATURAS E SIGLAS

CCM	Condição Causal de Markov
DAG	<i>Directed acyclic graph</i>
CPDAG	<i>Complete partial directed acyclic graph</i>
STROBE	<i>Strengthening the Reporting of Observational Studies in Epidemiology</i>

SUMÁRIO

	INTRODUÇÃO	12
1	OBJETIVOS	17
2	MODELOS CAUSAIS FORMAIS	18
2.1	Tipos de perguntas	21
2.2	Algumas premissas da inferência causal	26
3	COMUNICAÇÃO DE ESTUDOS EPIDEMIOLÓGICOS	39
3.1	Comunicação de fontes de incerteza	41
3.2	Possíveis diferenças na percepção e no relato de incertezas	49
3.2.1	<u>Revisão editorial e diretrizes de relato</u>	53
3.2.2	<u>Diferenças na atenuação da linguagem escrita</u>	55
4	APLICAÇÕES E CARACTERÍSTICAS DOS DIAGRAMAS CAUSAIS	57
4.1	Introdução	57
4.2	Métodos	57
4.2.1	<u>Seleção dos estudos</u>	57
4.2.2	<u>Coleta e análise de dados</u>	61
4.2.2.1	Análises bibliométricas	61
4.2.2.2	Características dos diagramas causais	63
4.2.2.3	Relato de confundimento	64
4.2.2.4	Análises de dados	66
4.3	Resultados	66
4.3.1	<u>Análises bibliométricas das publicações com diagramas causais</u>	66
4.3.2	<u>Características dos diagramas causais</u>	71
4.3.3	<u>Relato de confundimento</u>	73
4.4	Discussão	75
4.4.1	<u>Construção e apresentação do DAG causal</u>	77
4.4.2	<u>Relato de confundimento</u>	81
5	RELATO DAS LIMITAÇÕES DO ESTUDO E MARCADORES DE INCERTEZA	84
5.1	Introdução	84
5.2	Métodos	85
5.2.1	<u>Seleção dos estudos</u>	85

5.2.2	<u>Coleta de dados</u>	85
5.2.2.1	Relato das limitações do estudo	85
5.2.2.2	Marcadores linguísticos de incerteza.....	86
5.2.2.3	Características gerais das publicações e dos coautores	87
5.2.3	<u>Análises de dados</u>	88
5.3	Resultados	88
5.4	Discussão	92
5.4.1	<u>Interpretações da pergunta de pesquisa</u>	93
5.4.2	<u>Estruturas causais</u>	95
5.4.3	<u>Limitações</u>	101
	CONCLUSÃO	104
	REFERÊNCIAS	105
	APÊNDICE A - Estratégias de buscas	124
	APÊNDICE B - Questionário	127
	APÊNDICE C - Tesouro	137
	APÊNDICE D - Lista de pistas <i>Hedges</i>	146
	APÊNDICE E – Script em r.....	147

INTRODUÇÃO

O raciocínio causal é uma de nossas características cognitivas mais centrais. Na nossa vida cotidiana, tentamos frequentemente identificar quais acontecimentos conduziram a um determinado desfecho, gerando assim explicações para os eventos. Além disso, utilizamos o raciocínio causal para prever os efeitos de nossas ações ou quando simulamos mentalmente como poderíamos ter prevenido ou modificado um dado desfecho se algum aspecto da situação tivesse sido diferente. Esta capacidade de simular alternativas aos eventos que aconteceram no passado é chamada de pensamento contrafactual. Alguns autores sugerem que a ativação desse tipo de raciocínio ocorre, especialmente, em situações com uma resposta afetiva negativa, ligadas a uma evidente violação de nossas expectativas e para eventos de maior proximidade física ou temporal¹ (KAHNEMAN & TVERSKY, 1982; SENOS, 2008; VAN HOECK ET AL, 2015). Em todos os casos, o raciocínio causal parece resultar da necessidade de lidar com a incerteza, no sentido de repor a percepção de previsibilidade e controle sobre a realidade.

¹ Van Hoeck et al (2015) propõe tres estágios de processamento do raciocínio contrafactual: *ativação, inferência e adaptação*. Para ilustrar estas etapas os autores utilizam o exemplo de uma situação onde é recebida uma carta de rejeição de um novo emprego. O recebimento da carta levaria à ativação de memórias anteriores (p.ex. detalhes sobre o processo de apresentação do pedido e da entrevista) e provocaria uma simulação mental do evento (p.ex., considerando as exigências e expectativas do empregador). À medida que a ativação se espalha, ela aciona simulações mentais de situações semelhantes (aplicações anteriores), permitindo inferir outros desfechos para esse cenário se as circunstâncias fossem diferentes (p.ex., se o indivíduo tivesse dedicado mais tempo se preparando para a entrevista, ou seja, o evento contrafactual). Essas simulações contrafactuais (de eventos próximos) influenciam a interpretação do indivíduo sobre a experiência factual (p.ex., a recusa recebida foi causada pela incapacidade de se preparar adequadamente para a entrevista) e promovem um comportamento adaptativo que orienta o planejamento futuro e a resolução de problemas. Portanto, de acordo com Van Hoeck et al (2015), o processo de raciocínio contrafactual suporta inferências sobre o evento factual que orientam a compreensão da situação, modula as respostas emocionais e fornecem a base para o planejamento futuro e a tomada de decisão.

Deste modo, a busca por conhecimento causal tem sido associada a dois objetivos básicos (RUSSO & WILLIAMSON, 2007): o primeiro, que pode ser descrito como “objetivo cognitivo”, refere-se à tentativa de gerar explicações para os fenômenos na natureza e, assim, aumentar a nossa compreensão do mundo. O segundo objetivo que é “orientado para ação” visa, com base no conhecimento causal, realizar predições sobre eventos e decidir quais as melhores estratégias de intervenção.

A Epidemiologia moderna oferece um bom exemplo deste duplo propósito da inferência causal. Como ciência, a Epidemiologia utiliza o raciocínio causal na tentativa de compreender os fenômenos em saúde, por meio do estudo dos seus determinantes em grupos populacionais. Como disciplina da saúde pública, busca gerar conhecimentos que possam ser aplicados no desenvolvimento de estratégias voltadas à prevenção e à promoção da saúde em populações.

No entanto, essa perspectiva causal nem sempre é considerada de forma explícita entre os epidemiologistas. Como destacado em Russo & Williamson (2007), o vocabulário (causal) epidemiológico é, por vezes, substituído por termos genéricos, como fatores de risco ou determinantes, mas não “causa e efeito”. A não adoção de uma linguagem causal (explícita) pode estar associada às dificuldades metodológicas e práticas encontradas por epidemiologistas ao inferir relações causais a partir dos dados.

Além disso, Kincaid (2011) argumenta que as origens da epidemiologia estão fortemente ligadas à dúvida “positivista” acerca da legitimidade da causalidade como uma noção científica e cita a influência dos trabalhos de Karl Pearson na epidemiologia tradicional. Em *Grammar of Science* (1892), Pearson argumenta que o vocabulário metafísico da causalidade deveria ser substituído por uma linguagem com maior rigor matemático – ele destaca, por exemplo, que o conceito de “correlação” proporciona uma alternativa mais abrangente e mais flexível ao conceito clássico de causa (PEARSON, 1892).

Contudo, nos últimos anos têm se observado mudanças na forma com que a inferência causal é tratada na epidemiologia. Nesse contexto, destacam-se as novas formalizações causais para alguns conceitos epidemiológicos tradicionais, como a noção de confundimento e viés de seleção, e a adoção de novos conceitos causais como colisores, *M-bias*, caminhos pela porta de trás, confundidores tempo-dependentes, positividade, transportabilidade, etc.

Essas mudanças tiveram início com o desenvolvimento do modelo de respostas potenciais (NEYMAN, 1923; RUBIN, 1975; 1976; 1990; HOLLAND, 1986) e, mais recentemente, da teoria dos diagramas causais (PEARL, 1995; 2000; GREENLAND, ET AL 1999). De forma geral, os diagramas causais são gráficos acíclicos direcionados (DAG) que codificam as hipóteses qualitativas do pesquisador sobre o processo de geração dos dados. A teoria dos diagramas causais pode ser aplicada a uma série de questões na pesquisa epidemiológica, como a identificação de parâmetros causais e aspectos relacionados à validade externa e a síntese dos resultados (PEARL, 2010; 2012; PEARL & BAREINBOIM, 2014). Os diagramas causais também têm sido utilizados para demonstrar em quais situações o uso de critérios e técnicas tradicionais para tratar confundimento e viés de seleção podem levar a conclusões inadequadas (HERNÁN et al., 2002; 2004; ROBINS, 2001). Além disso, permitiram compreender algumas controvérsias presentes na literatura epidemiológica, como a natureza do viés de Berkson e do Paradoxo de Simpson (PEARL, 2014a).

Esse cenário parece caracterizar, para alguns autores, o surgimento de uma “revolução metodológica” na área (PORTA & BOLÚMAR, 2015; BROADBENT, 2015). Porta & Bolúmar (2015) argumenta que “essas inovações conceituais e metodológicas podem estar modificando a forma com que a pesquisa clínica e epidemiológica é concebida e realizada, e como avaliamos a validade e relevância dos resultados”. Ele acrescenta que essas mudanças em curso podem ter um grande impacto sobre a produção científica, mas que esse impacto ainda não foi quantificado.

A avaliação dos efeitos da utilização de ferramentas de inferência, como os diagramas causais, pode ser feita de várias maneiras. O impacto de uma ferramenta pode ser medido examinando o seu grau de disseminação ou avaliando como os seus pressupostos e regras são reconhecidos e interpretados nos estudos (PENFIELD et al, 2014). Neste contexto, destaca-se que as conclusões inferidas a partir de um modelo causal estrutural dependem da hipótese de que a estrutura gráfica descreve corretamente os processos de geração dos dados. Considerando que algumas dificuldades, tais como confundidores não mensurados, dados faltantes e erros de mensuração, são comuns em muitas áreas de pesquisa, alguns autores têm reforçado a importância da construção de modelos causais “realistas” que representam simultaneamente as fontes comuns de incerteza em estudos

epidemiológicos (GREENLAND, 2010; HERNÁN & ROBINS, 2016; HERNÁN et al, 2012). Deste modo, outro elemento que pode ser investigado é a frequência com que esses aspectos relacionados às limitações do estudo são incluídos nos DAG ou como são comunicadas as possíveis fontes de incerteza sobre a estrutura causal.

Além disso, uma das principais hipóteses apresentadas anteriormente é que a utilização de diagramas causais pode contribuir para a qualidade metodológica e do relato de estudos observacionais (PETERSEN & VAN DER LANN, 2014; PORTA & BOLÚMAR, 2015; GLYMOUR, 2006). A construção de modelos causais durante o planejamento das análises ou durante a redação do artigo pode contribuir, por exemplo, para uma avaliação mais completa das fontes de viés que podem ter afetado os resultados do estudo. No entanto, também há uma preocupação de que, dependendo da maneira como são utilizadas, essas ferramentas poderiam colaborar para inferências excessivamente confiantes (GREENLAND, 2010; 2017a; DAWID, 2010).

No artigo "*Be aware of the DAG!*" - publicado em 2010, Dawid argumenta que a inferência causal em estudos observacionais requer um conjunto de hipóteses nem sempre passíveis de serem testadas por meio dos dados e, portanto, a justificativa prévia dessas hipóteses é um fator fundamental para utilização desses modelos gráficos como modelos causais (DAWID, 2010).

Em "*Overthrowing the tyranny of null hypotheses hidden in causal diagrams*", Greenland (2010) argumenta que, dado as limitações dos dados e a complexidade dos sistemas causais comumente estudados na epidemiologia, grande parte dos modelos causais, como aqueles apresentados como gráficos acíclicos direcionados (DAG), podem ser considerados representações pouco realistas para os temas de pesquisa. Como destacado em Greenland (2010), a razão para esta afirmação é que, em geral, "a identificação demanda a abstração de aspectos importantes do modelo, impondo, assim, hipóteses nulas que não se justificam, tanto em observações passadas como no conhecimento a priori.

Uma maneira de abordar essa questão é avaliar se há maior ou menor autocrítica (reconhecimento de mais limitações do estudo ou fontes de viés) e melhor avaliação dessas limitações entre artigos com e sem a utilização de DAG causais. No entanto, a comunicação de incerteza também pode ser feita de forma "implícita", por meio de recursos linguísticos. Ou seja, a incerteza pode ser comunicada não apenas pela presença de certas informações no texto, mas também

pela maneira na qual os autores expressam essas informações. No contexto linguístico, o uso de *hedges* tem sido um dos principais aspectos investigados e pode ser descritos como recursos linguísticos que expressam a extensão da incerteza dos autores sobre uma determinada sentença (HYLAND, 2004). Deste modo, o estudo dos termos *hedges*, em conjunto com os demais aspectos do relato, pode contribuir para avaliação mais completa da comunicação de fontes de incerteza em estudos observacionais.

1 OBJETIVOS

Neste trabalho, foi realizada uma revisão da literatura com o objetivo de descrever o quanto, como e onde os diagramas causais têm sido utilizados nos estudos observacionais analíticos nos últimos 18 anos. Foram analisados os itens relativos ao crescimento anual das publicações, os principais periódicos, os padrões de autoria e coocorrência de palavras chaves.

O segundo objetivo foi descrever as principais aplicações e características dos modelos gráficos empregados nessas publicações. Os principais itens investigados foram: (i) a frequência com que os aspectos relacionados às limitações do estudo são incluídos nos DAG; (ii) a comunicação de possíveis fontes de incerteza sobre a estrutura causal e (iii) e a frequência do relato adequado de itens importantes na comunicação de confundimento em estudos observacionais.

Em um segundo estudo, duas amostras de artigos com e sem diagramas causais foram comparadas quanto ao relato das limitações do estudo e o uso de marcadores linguísticos de incerteza. As análises apresentadas são utilizadas para discutir algumas dificuldades metodológicas relacionadas à formulação da pergunta de pesquisa sob uma perspectiva causal e a construção de diagramas causais.

2 MODELOS CAUSAIS FORMAIS

Como destacado em Pearl (2000), a teoria estrutural da causalidade pressupõe que o mundo está organizado sob a forma de *mecanismos estáveis* ou leis físicas. Além disso, esses mecanismos são *autônomos* – como ligações mecânicas em uma máquina – no sentido de que podemos realizar modificações em alguns mecanismos sem alterar os demais. Nesse contexto, as intervenções sempre envolvem a quebra de algum mecanismo e a causalidade nos diz qual mecanismo deve ser modificado.

Deste modo, na teoria estrutural, a causalidade é vista como um resumo de comportamentos sob intervenções e a inferência causal como uma atividade que visa prever, com base em um conjunto de premissas, os efeitos de certas intervenções. Além disso, os DAG (ou as equações estruturais) são utilizados como uma linguagem matemática na qual as premissas causais podem ser formuladas e manipuladas. Por fim, as intervenções são tratadas como cirurgias atômicas em equações ou no gráfico.

As noções causais de autonomia (ou modularidade) e direcionalidade são codificadas em um modelo chamado de *modelo causal estrutural*. Como destacado em Sloman (2005) esse modelo se relaciona com três entidades principais:

1. **O sistema causal de interesse:** composto por um conjunto de mecanismos (observados e não observados)
2. **Um conjunto de dados empíricos:** representado por uma distribuição de probabilidade para as variáveis observadas.
3. **Um DAG ou equações estruturais:** que representa as hipóteses causais sobre os processos que geraram os dados.

Deste modo, precisamos determinar como o DAG se relaciona com a parte da natureza que está sendo representada. Isto é, como podemos codificar no gráfico as hipóteses causais sobre o fenômeno de interesse – o que devemos incluir no gráfico e o que pode ser abstraído, tendo em vista os objetivos da inferência causal. Além disso, precisamos determinar como as hipóteses causais codificadas no DAG se relacionam com os dados observados. Em outras palavras, como podemos avaliar a consistência entre o que sabemos, o que foi observado e a pergunta que queremos responder.

De forma geral, o sistema de interesse (parte da natureza que está sendo representada no modelo) é descrito por variáveis aleatórias, que podem ser classificadas em dois conjuntos. Um conjunto de variáveis endógenas (V) e um conjunto de variáveis exógenas (U). Além disso, o modelo é composto por um conjunto de funções F . Cada função em F representa um mecanismo distinto que descreve como cada variável em V obtém os seus valores no sistema. Considere as seguintes funções:

$$\begin{aligned}z &= f_Z(u_Z) \\x &= f_X(z, u_X) \\y &= f_Y(x, u_Y)\end{aligned}$$

Essas funções são ditas estruturais no sentido de serem autônomas, ou seja, se pudermos assumir que as mudanças externas que afetam uma função, não implicam em alterações nas relações (mecanismos) descritas nas demais funções. Deste modo, um modelo causal descreve uma série de experimentos hipotéticos e qualquer subconjunto de funções estruturais é, em si mesmo, um modelo “válido” da realidade – um que prevalece sob algum conjunto de intervenções (PEARL, 2010, p.49).

Cada uma dessas funções representa o processo causal (ou mecanismo) que determina o valor da variável do lado esquerdo da equação a partir dos valores das variáveis do lado direito (PEARL, 2000). Neste contexto, uma variável X é uma causa direta de uma variável Y se X aparece na função (f_Y) que atribui o valor de Y . Além disso, X é uma causa de Y se X é uma causa direta de Y , ou de outra causa de Y (PEARL ET AL, 2016). Por exemplo, nas equações apresentadas, X é uma causa direta de Y , e Z é uma causa não-direta de Y . As variáveis exógenas U_Z, U_X, U_Y , representam as condições de fundo (background factors) para a qual nenhum mecanismo explicativo é codificado no modelo, mas que influenciam as demais variáveis (endógenas) no modelo.

Um diagrama causal é definido a partir de um modelo causal estrutural da seguinte forma (SHPIETSER, 2008): cada variável em V corresponde a um vértice no gráfico. Além disso, uma variável X (contida no conjunto V ou U) e outra variável Y estarão ligadas por uma seta direcionada de X a Y , se X aparece na função (f_Y) que atribui o valor de Y .

A ausência de uma variável do lado direito da equação codifica o pressuposto de que essa variável não participa diretamente no processo de determinação do valor da variável de saída. Por exemplo, na função $y = f_Y(x, u_Y)$, a ausência da variável Z codifica a hipótese de que as mudanças externas que afetam Z não implicarão em variações em Y , dado que os X, U_Y permaneçam constantes (PEARL, 2010). Essa hipótese mais definitiva de efeito causal direto nulo é representada graficamente pela ausência de uma seta de Z até Y .

Na prática, dificilmente temos acesso a todas as causas que afetam diretamente uma determinada variável a ponto de poder estabelecer uma equação que descreva corretamente como uma variável obtém o seu valor na natureza (SHPITSER, 2008). Deste modo, a nossa ignorância sobre o contexto de fundo (ou seja, tudo que é deixado de fora do modelo) é representado por uma distribuição de probabilidades $P(u)$. Esta distribuição, em conjunto com os mecanismos (as funções) que descrevem as causas diretas de cada variável em V , induz uma distribuição de probabilidades $P(v)$ para as variáveis do domínio observável (SHPITSER, 2008). Ao assumir que as variáveis que foram deixadas de fora do modelo são mutuamente independentes em $P(u)$ ou seja, que as variáveis em U não compartilham uma causa em comum e, portanto, aparecem no máximo em duas funções, então podemos omitir essas variáveis do gráfico.

Além disso, se o gráfico associado ao modelo não contém ciclos, ou seja, pode ser representado por um DAG, então esse modelo, chamado de recursivo, terá um conjunto de propriedades úteis que nos permite relacionar a distribuição de probabilidades $P(v)$ com a estrutura gráfica. O gráfico definido desta forma, a partir de um modelo causal estrutural, é dito induzido por M , e chamado de diagrama causal (PEARL, 2010). Essa relação entre um conjunto de distribuição de probabilidades e o modelo gráfico é dada pela Condição Causal de Markov (CCM).

Condição Causal de Markov (PEARL, 2010)²: qualquer distribuição de probabilidades gerada por um modelo markoviano M pode ser fatorada como:

$$P(v_1, v_2, \dots, v_n) = \prod_i P(v_i | pa_i), \quad (1)$$

² Este teorema foi apresentado pela primeira vez em Pearl e Verma (1991). No entanto, como destacado em Pearl (2010), a ideia geral da CCM está implícita nas obras de Kiiveri et al (1984).

onde V_1, V_2, \dots, V_n são as variáveis endógenas em M , e pa_i são os (valores dos) pais (causas diretas endógenas) de V_i no diagrama causal associado a M .

Basicamente, a CCM nos diz que nas distribuições de probabilidade compatíveis com a estrutura gráfica, cada variável representada no DAG é independente das demais variáveis (exceto os seus efeitos), dado a informação sobre as suas causas diretas (pais). Uma descrição alternativa é que, para especificar completamente a distribuição conjunta (ou seja, a probabilidade de qualquer combinação particular dessas variáveis representadas no gráfico) e, portanto, o modelo estatístico, precisamos considerar apenas a distribuição condicional de cada variável dado os seus pais. Deste modo, a tarefa de descrever um modelo complexo de probabilidade conjunta para um problema específico é simplificada pela especificação de uma série de relações "locais" entre pais e filhas (entre cada variável e suas causas diretas).

2.1 Tipos de perguntas

Diferentes perguntas, tais como as relativas à associação entre variáveis, aos efeitos de intervenções ou a situações contrafactuais, são representadas como contrastes entre diferentes distribuições de probabilidade para as variáveis de interesse (SHPITSER, 2008). Pearl (2010) propõe uma classificação para essas perguntas na forma de uma hierarquia de 3 níveis. Essa hierarquia baseia-se nas diferenças em relação às informações necessárias para se obter uma resposta em cada nível.

Tabela 1. Hierarquia de perguntas

Nível	Atividade geral	Tipo de pergunta
Associação $Pr(y x)$	Observar (<i>seeing</i>)	Qual é? Ao observar X eu mudaria a minha crença sobre Y?
Intervenção $Pr(y do(x))$	Fazer/Intervir (<i>doing</i>)	O que acontece se? O que acontece se eu faço X?

Contrafactuais $Pr(y_x x', y')$	Imaginar Retrospecto (Imagining)	Por quê? X foi a causa de Y? E se eu tivesse agido de forma diferente?
---	--	--

Fonte: PEARL & BAREINBOIM, 2014.

No primeiro nível encontram-se as perguntas relacionadas à associação entre variáveis. Como destacado em Shpitser (2008), tais perguntas são representadas como distribuições de probabilidade marginal ou condicional e podem ser respondidas a partir da distribuição conjunta sobre as variáveis do domínio de interesse. Alguns exemplos de perguntas desse tipo são: a prevalência de doenças cardiovasculares entre fumantes é maior do que entre não fumantes?

A probabilidade de se observar um evento (por exemplo, a ocorrência de doença cardiovascular) em algumas circunstâncias (entre indivíduos fumantes) pode ser representada como a probabilidade condicional de uma variável aleatória Y em algum nível “y” quando X é observada a ter algum valor “x”. A probabilidade condicional de Y dado X, com $P(X) > 0$, pode ser definida como:

$$P(Y = y | X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)} \quad (2)$$

De forma geral, o objetivo da análise estatística padrão é avaliar os parâmetros de uma distribuição populacional a partir de amostras extraídas dessa população. Com a ajuda de tais parâmetros, busca-se inferir associações entre variáveis, estimar probabilidades de eventos, bem como atualizar essas probabilidades de acordo com novas evidências (PEARL, 2000).

No segundo nível encontram-se as perguntas relacionadas à predição dos efeitos de intervenções. Alguns exemplos de perguntas desse tipo são: qual o efeito da cessação do tabagismo sobre a ocorrência de doenças cardiovasculares?

Como descrito em Pearl (2000), essas perguntas são de natureza causal, pois requerem algum conhecimento sobre os processos que deram origem os dados e, por essa razão, não podem ser respondidas a partir de premissas puramente probabilísticas. A análise causal requer uma etapa adicional à análise estatística; o

seu objetivo é estimar probabilidades de certos eventos, não somente, sob condições estáticas, mas também sob condições mutáveis, induzidas por intervenções externas.

Como descrito em Pearl (2000), não há nada na função de distribuição de probabilidades que nos permite inferir o comportamento dessa distribuição se as condições externas fossem diferentes. Deste modo, o comportamento probabilístico do sistema original (na ausência de intervenção) não pode ser traduzido diretamente para o comportamento probabilístico do sistema pós-intervenção (SHPITSER, 2008).

Essa distinção é descrita por alguns autores como a diferença entre observar (*seeing*) e intervir (*doing*) (PEARL, 2000; SPIRITES ET AL, 1993). “*Seeing*” envolve a observação passiva de um sistema em seu estado natural, enquanto “*doing*” relaciona-se com o comportamento do sistema em um estado perturbado, geralmente provocado por alguma intervenção externa.

Uma intervenção é definida em Pearl (2000) como uma “cirurgia atômica” que remove inteiramente os mecanismos que determinam uma variável, enquanto preserva as demais relações no modelo. Isso é, uma intervenção imposta a uma variável X é aquela que interrompe completamente o relacionamento entre X e seus pais (causas diretas) e define um novo valor constante ($X = x$). Além disso, essa intervenção é “cirúrgica”, pois não altera as demais relações causais no modelo (e distribuições de probabilidade condicionais correspondentes). Na linguagem gráfica, esse tipo de intervenção é chamado de intervenção ideal e é denotado por *do* ($X = x$). Em outras palavras, uma intervenção ideal em X pode ser vista como uma variável exógena (como um fator externo ao modelo) que determina totalmente o valor de X e que age diretamente apenas sobre X , permitindo que as mudanças nas demais variáveis sejam produzidas (ou não) por essa manipulação. Ou seja, as intervenções ideais são perfeitamente seletivas nas variáveis nas quais elas atuam diretamente.

Pearl (2000) sugere a notação $\Pr(Y \mid \text{do}(X = x))$ e $\Pr(Y \mid X = x)$ para distinguir entre a distribuição de probabilidade condicional dada uma intervenção em X e dada a observação passiva de X , respectivamente. O primeiro reflete como a distribuição de Y deve ser modificada, dada a informação de que X tenha sido manipulado, por alguma intervenção externa, a assumir o valor “ x ”, enquanto a segunda reflete como a distribuição de Y deve ser modificada, dado que se verifica que $X = x$.

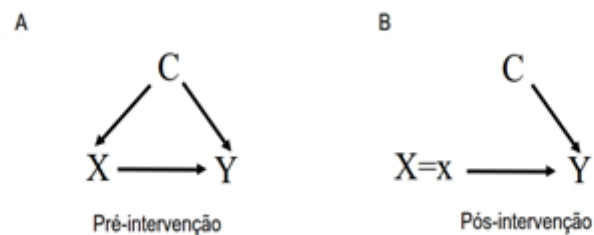
Por exemplo, para dois conjuntos disjuntos de variáveis X e Y , o efeito causal de X em Y é dado pela distribuição de probabilidade de Y quando uma intervenção do tipo $X = x$ é aplicada de maneira uniforme sobre a população. Em geral, podemos definir formalmente a distribuição de probabilidade pós-intervenção pela equação (PEARL, 2010):

$$P_M(y | do(x)) \triangleq P_{M_x}(y) \quad (3)$$

Deste modo, os efeitos causais são definidos em termos processuais, por meio de simulações de intervenções em um modelo causal M , resultando em um submodelo M_x . No âmbito do modelo M , a distribuição de probabilidade pós-intervenção para o desfecho Y é definida como a probabilidade que o modelo M_x atribui a cada nível do desfecho $Y = y$ (PEARL, 2010).

Na figura 1 é apresentado um diagrama causal para um modelo M e o gráfico que representa a simulação de uma intervenção $do(x)$ (isto é, o gráfico induzido do submodelo M_x). Nota-se que o objetivo principal é expressar as distribuições na forma $P(Y=y | do(X=x))$ em termos da distribuição conjunta para as variáveis observadas.

Figura 1 Exemplo de uma intervenção $do(x)$ em um DAG causal.



Fonte: CORTES, 2018.

Na seção anterior, a relação entre um DAG causal e a distribuição de probabilidade sobre as suas variáveis foi descrita pela Condição causal de Markov. Deste modo, a distribuição conjunta pré-intervenção, regida pelo modelo da Figura 2A, é dada pelo produto: $P(C = c, X = x, Y = y) = P(C = c) P(X = x | C = c) P(Y = y | X = x, C = c)$. Dado que a intervenção ideal em X interrompe completamente o

relacionamento entre X e seus pais e define um novo valor constante “ x ”, então o fator $P(X=x | C=c)$ é removido do produto acima. Além disso, em um modelo causal estrutural, as relações entre uma variável e os seus pais são vistas como mecanismos autônomos que podem ser manipulados de forma independente, sem afetar as demais relações do modelo (e as distribuições de probabilidade condicionais correspondentes). Em outras palavras, as noções de intervenção ideal e autonomia (ou modularidade) assumem que as variáveis afetadas pela intervenção não são mais influenciadas pelo seu conjunto de pais e as demais variáveis do modelo são governadas pelos processos pré-intervenção. Deste modo, a distribuição conjunta pós-intervenção, regida pelo modelo da Figura 1B, é dada pelo produto: $P(C=c, Y=y | do(X=x)) = P(C=c) P(Y=y | X=x, C=c)$. A partir desta distribuição, o efeito causal de X em Y pode ser obtido pela equação:

$$P(Y = y | do(X = x)) = \sum_c P(c)P(y|c, x) \quad (4)$$

No entanto, nem todas as perguntas de caráter causal podem ser codificadas por expressões do tipo “ $P(Y | do(x))$ ”, refletindo o fato de que nem todas as perguntas causais podem ser respondidas a partir de estudos experimentais (PEARL, 2010). No terceiro nível da hierarquia encontram-se as perguntas do tipo: qual é a fração dos casos de morte devida a um tratamento específico?

Para responder a tais questões é necessária uma análise probabilística de expressões contrafactuais do tipo “ Y teria tido o valor y , se X tivesse o valor x em uma situação de $U = u$ ” - denotadas por $Yx(u)y$

Essas expressões são ditas contrafactuais, porque lidam com quantidades não observadas, que diferem das quantidades observadas. Raciocinar causalmente sobre como as coisas poderiam ter sido diferentes requer a alteração de um modelo do mundo atual e, a partir de alguma suposição contrafactual, cria-se um modelo de um mundo diferente. Desta forma, podemos tratar a suposição contrafactual como uma intervenção imaginada. Para isso, alguma variável no modelo do “mundo atual” deve ser ajustada para o valor especificado pela intervenção imaginária.

Ou seja, a ideia principal é interpretar a expressão “se X tivesse o valor x ” como uma instrução para fazer uma modificação mínima do modelo atual, atribuindo a X um valor diferente do observado (PEARL, 2010). Deste modo, questões

contrafactuais (no nível individual) do tipo “O valor de Y na unidade u, se X tivesse o valor x” podem ser formalizadas da seguinte maneira (PEARL, 2000):

$$Y_x(u) \triangleq Y_{M_x}(u), \quad (5)$$

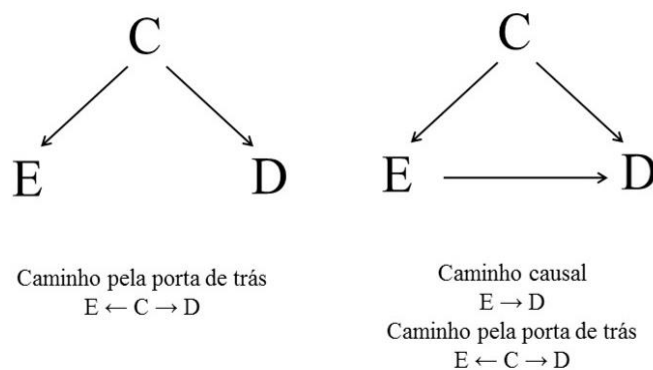
M é um modelo causal estrutural (atual), M_x é a versão modificada de M – que representa a simulação de uma intervenção *do* ($X=x$). $Y_{M_x}(u)$ denota a resposta para Y dada por M_x . Nesta abordagem, a unidade experimental u é igualada a um vetor que representa as condições de fundo $U = u$. Como descrito em Pearl (2010), isso reflete o entendimento de que os fatores importantes que caracterizam uma unidade (p.ex., um indivíduo) são os aspectos que determinam o seu comportamento ou resposta. Ou seja, a ideia é que as diferenças individuais resultam apenas das diferenças de seus atributos $U = u$, mas as leis da natureza, representadas nas funções f_x , f_y , etc, permanecem invariantes (PEARL, 2010).

Nota-se, que nas expressões contrafactuais, para cada situação $U=u$, obtemos um valor definido, $Y_x(u)$, que representa o valor hipotético de Y nessa situação. Por outro lado, as expressões de intervenção, denotadas pelo operador *do*(.), são definidas apenas em distribuições de probabilidade e sempre fornecem resultados probabilísticos (PEARL et al, 2016).

2.2 Algumas premissas da inferência causal

Confundimento é uma das preocupações centrais na investigação epidemiológica. De forma geral, o confundimento ocorre quando há uma fonte adicional de associação entre a exposição e o desfecho, que resulta da presença de causas comuns entre essas variáveis. Na linguagem gráfica, essa fonte adicional de associação se traduz na presença de caminhos pela porta de trás (desbloqueados) que ligam a exposição ao desfecho (Figura 2).

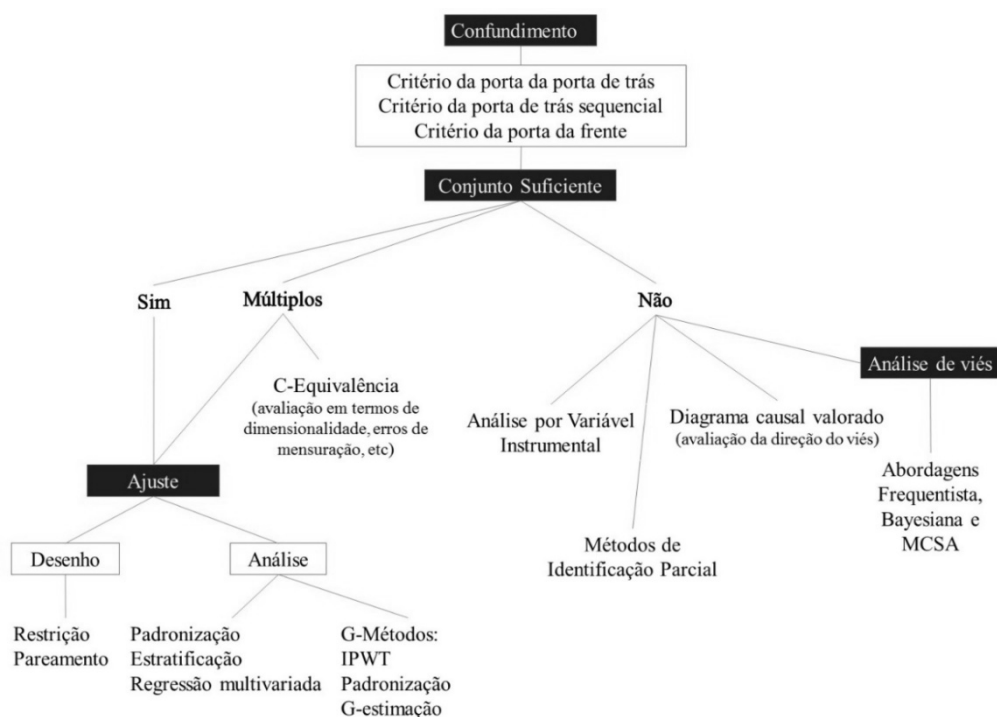
Figura 2. Representação estrutural de confundimento.



Fonte: CORTES, 2018.

Deste modo, uma estratégia para o controle de confundimento é selecionar um conjunto de covariáveis capaz de bloquear todos os caminhos pela porta de trás, enquanto mantém desbloqueados todos os caminhos causais que ligam a exposição ao desfecho. A figura 3 apresenta algumas estratégias de seleção e ajuste para esse conjunto de covariáveis.

Figura 3 Estratégias de controle de confundimento.



Fonte: CORTES, 2018.

No entanto, como destacado em Szklo (2015), a informação sobre uma associação entre duas variáveis é útil mesmo na presença de confundimento, na medida em que permite a identificação de grupos de risco elevado, permitindo que programas de rastreamento sejam implementados. Por exemplo, considere a associação estatística entre o hábito de fumar (A), a presença de dedos amarelados (B) e ocorrência de câncer de pulmão (D). Neste contexto, dado o conhecimento sobre a probabilidade de B é possível prever a probabilidade de D. No entanto, apenas o conhecimento causal sobre as relações entre essas variáveis permitirá prever que uma intervenção em B (por exemplo, a utilização de um produto químico para clarear os dedos) não terá um efeito sobre a ocorrência de câncer de pulmão (D), enquanto uma estratégia de prevenção do tabagismo (A) provavelmente terá um efeito na diminuição da incidência de D.

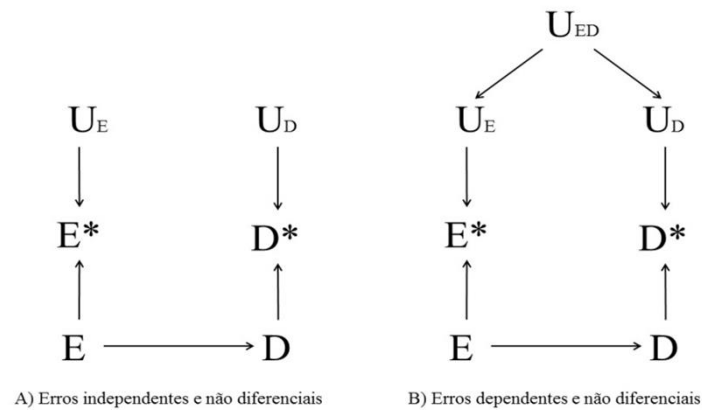
Neste caso, a associação entre B (dedos amarelados) e D (câncer de pulmão) é explicada pelo hábito de fumar (A). Neste exemplo trivial, o rastreamento de indivíduos com dedos amarelados permite a identificação do grupo (fumantes crônicos) que apresenta um risco elevado para ocorrência de câncer de pulmão. No entanto, se o objetivo é o desenvolvimento de estratégias voltadas à prevenção (intervenções em D), então é necessário investigar os mecanismos causais que influenciam o desfecho. Deste modo, o conhecimento sobre a estrutura causal é crucial para a inferência causal.

Além da incerteza sobre o conjunto suficiente de confundidores, outra situação que pode implicar em confundimento (residual) é a presença de erros de mensuração das covariáveis selecionadas. Muitas exposições e desfechos de interesse em epidemiologia são passíveis de erros de mensuração.

A ocorrência de erros de classificação tanto da exposição e desfecho quanto das covariáveis pode resultar em distorções nas estimativas de efeito causal obtidas no estudo. Essas questões tornam-se ainda mais complexas quando as relações causais de interesse envolvem variáveis que não são diretamente observáveis. Essas variáveis (latentes) necessitam, portanto, de uma conceituação que lhes dê operacionalidade, e a mensuração dá-se através de medidas indicadas pela definição operacional (indicadores). Deste modo, a má classificação pode resultar tanto de problemas operacionais no processo de redução de conceitos a variáveis, quanto de problemas na mensuração em si (REICHENHEIM & MORAES, 1998).

Pode-se distinguir quatro tipos de erros de mensuração (HERNÁN et al, 2009): (1) erros independentes e não diferenciais; (2) erros dependentes e não diferenciais; (3) erros independentes e diferenciais e (4) erros dependentes e diferenciais.

Figura 4. Representação estrutural de erros de mensuração.



Fonte: HERNÁN et al, 2009.

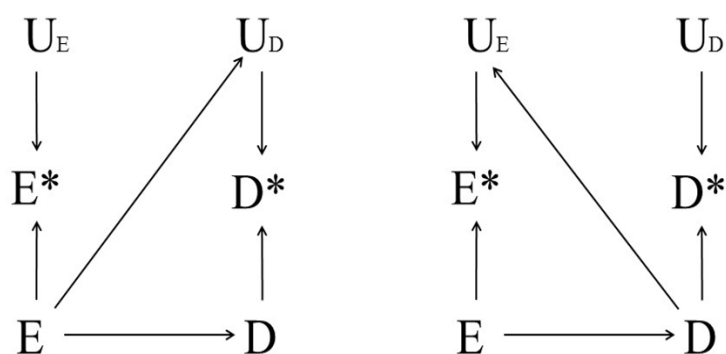
A Figura 4A mostra a estrutura geral para uma situação de erros de mensuração independentes e não diferenciais. Considere um estudo cujo o objetivo é avaliar o efeito de uma exposição E sobre um desfecho D (representado pela seta de E até D). Suponha que a exposição e o desfecho foram mensurados com erros, representado pelas variáveis substitutas E* e D*, respectivamente. A seta de E para E* traduz a ideia de que o status da exposição mensurada no estudo é influenciado pelo verdadeiro valor da exposição. As variáveis U_E e U_D representam os demais fatores que influenciam os valores de E* e D*.

O gráfico da figura 4A representa uma situação na qual os erros de mensuração (U_E e U_D) não compartilham causas em comum. Neste exemplo, ao aplicar as regras de d-separação, observamos que o único caminho entre os erros de mensuração (U_E → E* ← E → D → D* ← U_D) está bloqueado devido à presença de colisores (E* e D*). A Figura 4B apresenta uma situação de dependência entre os erros de mensuração da exposição e do desfecho. Destaca-se que, ao contrário da situação de confundimento, o viés de mensuração não é descrito por meio de caminhos (pela porta de trás) abertos entre a exposição verdadeira (E) e o desfecho verdadeiro (D); em vez disso, o viés de mensuração resulta da presença de

caminhos pela porta de trás entre a exposição (E^*) e o desfecho (D^*) mensurados com erros.

As figuras 4A e 4B representam configurações em que o erro de mensuração da exposição (U_E) é independente do valor real do desfecho, e o erro de mensuração do desfecho é independente do valor real da exposição. Ou seja, dizemos que os erros de mensuração da exposição são não diferenciais em relação ao desfecho, se a mensuração da exposição não é influenciada pelo verdadeiro valor do desfecho. Da mesma forma, dizemos que os erros de mensuração do desfecho são não diferenciais em relação à exposição, se a mensuração do desfecho não é influenciada pelo verdadeiro valor da exposição

Figura 5. Representação estrutural de erros de mensuração.

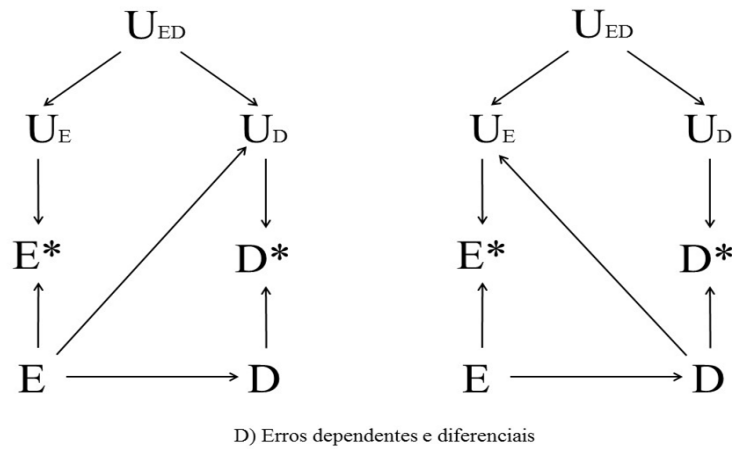


C) Erros independentes e diferenciais

Fonte: HERNÁN et al, 2009.

A figura 5 apresenta configurações de erros de mensuração independentes, mas diferenciais. No diagrama A o caminho $E \rightarrow U_D \rightarrow D^*$ indica que a mensuração do desfecho é influenciada pelo verdadeiro valor da exposição. No diagrama B, o caminho $D \rightarrow U_E \rightarrow E^*$ indica que a mensuração da exposição é influenciada pelo verdadeiro valor do desfecho. A figura 6 apresenta configurações de erros de mensuração não independentes e não diferenciais.

Figura 6. Representação estrutural de erros de mensuração.



Fonte: HERNÁN et al, 2009.

Uma heurística bem conhecida na literatura é que os erros de mensuração independentes e não diferenciais influenciam os valores esperados de um estimador em direção ao valor nulo. Como descrito em Jurek et al (2005) existem várias versões desta regra na literatura. Por exemplo, Bross (1954) argumentou que, considerando uma exposição dicotômica, os erros de classificação não diferenciais e independentes influenciam o estimador do risco relativo em direção ao valor nulo. Analogamente, Greenland (1980) argumentou que o ajuste por um fator de confusão com erro de classificação não diferencial produz uma medida parcial que pode variar entre o valor bruto (isto é, sem ajustes) e o valor verdadeiro (isto é, ajustado pelo verdadeiro fator de confusão). Além disso, Brenner (1993) demonstrou que o ajuste por um fator de confusão politômico com erros de mensuração não diferenciais pode induzir a um viés maior em comparação com a medida não ajustada pelo confundidor.

No entanto, tais regras são muitas vezes inaplicáveis e depende de outras condições além da não diferencialidade (e independência). Como destacado em Greenland & Gustafson (2006), entre os problemas bem conhecidos com tais argumentos é que a independência e a não diferencialidade são condições muito difíceis de satisfazer (por exemplo, por causa da correlação entre os erros de alguns itens de um questionário) e pequenas violações de independência ou não diferencialidade podem implicar em viés substancial. Além disso, no caso de exposições politômicas outras condições são necessárias para garantir a regra de

viés em direção ao valor nulo. A regra de ajuste parcial para confundidores também requer condições adicionais no caso de variáveis dicotômicas e politômicas (OGBURN & VANDERWEELE, 2012).

Ressalta-se que essas regras assumem a ausência de interações com outras fontes de erros sistemático, como viés de seleção e confundimento. Deste modo, como destacado em Jurek et al (2005), na prática, é difícil garantir que todas estas condições estejam satisfeitas, e práticas comuns podem levar a violações desses pressupostos. Por exemplo, no caso de uma exposição contínua ou politômica com erros não-diferenciais, a classificação ou recodificação dessas variáveis em algumas categorias pode resultar em erros diferenciais.

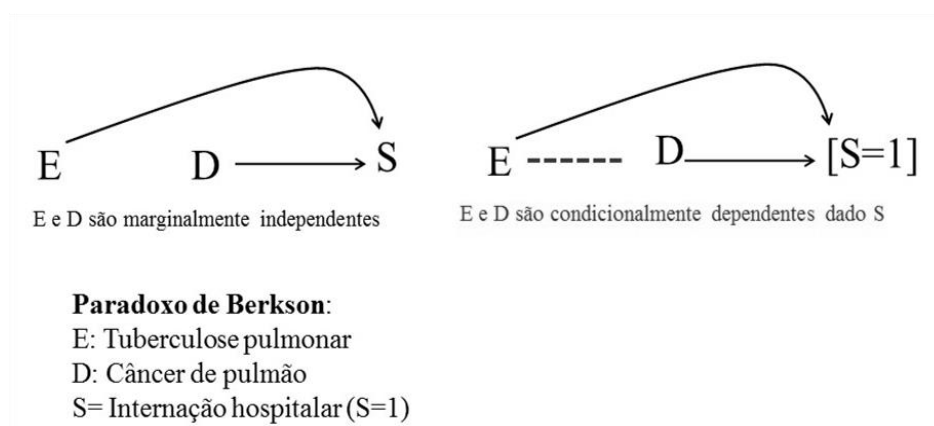
Além disso, as alegações sobre viés referem-se ao método que gerou uma estimativa (o estimador) e não a estimativa em si. Como destacado em Jurek et al (2006), essas regras referem-se a valores esperados de estimadores – isto é, o resultado médio da aplicação de uma fórmula (estimador) em replicações da amostra – e não a uma estimativa de um estudo específico. Deste modo, uma tendência do viés em direção ao valor nulo é insuficiente para garantir que a estimativa observada estará subestimada. Métodos gráficos para a identificação de efeitos causais na presença de erros de mensuração são apresentados em Kuroki & Pearl (2014).

Outro problema comum para inferência causal é a possibilidade de viés de seleção. De forma geral, o viés de seleção ocorre quando há uma fonte adicional de associação entre a exposição e desfecho, que resulta da inclusão seletiva de certos indivíduos para compor o conjunto de dados (BAREINBOIM & PEARL, 2014). Um dos parâmetros causais de interesse na epidemiologia é o efeito causal médio, isto é, a média dos efeitos causais individuais na população alvo. Deste modo, a população alvo inclui todos os indivíduos para os quais desejamos que as conclusões causais obtidas no estudo sejam válidas, enquanto a população da amostra inclui os indivíduos (selecionados a partir da população fonte) sobre os quais coletam-se os dados. Deste modo, a seleção dos indivíduos para a amostra (ou para a análise dos dados) pode ser indicada por uma variável dicotômica S ($S = \text{sim}$ ou $S = \text{não}$), e as inferências (para a população alvo) a partir da amostra são necessariamente condicionadas para o valor da variável S ($S = \text{sim}$). Idealmente, a população fonte é igual à população alvo, e os indivíduos são amostrados aleatoriamente a partir dessa população. No entanto, em geral, o viés de seleção

ocorrerá se a seleção dos indivíduos para compor o conjunto dos dados é influenciada pelo status da exposição e do desfecho (ou pelas causas não mensuradas de E e D). Como consequência, a estimativa do efeito causal médio (ECM) na população de estudo pode ser diferente do ECM na população alvo.

Em 1946, Joseph Berkson observou que duas doenças (E e D) que não estavam associadas na população (ou seja, eram marginalmente independentes) poderiam estar associadas entre pacientes hospitalizados, se ambas as doenças afetassem a probabilidade de admissão hospitalar. Como destacado na figura 7, a variável S indica que o conjunto de dados foi restrito aos indivíduos com internação hospitalar (BERKSON, 1946). Como consequência, E e D que eram independentes na população geral, estão associadas entre pacientes hospitalizados, que compõem o estrato $S = 1$.

Figura 7. Exemplo de viés de colisão (paradoxo de Berkson).



Fonte: CORTES, 2018.

Atualmente o viés (ou paradoxo) de Berkson é descrito como um tipo de viés de colisão (*collider* ou *colling bias*). Destaca-se que este tipo de viés pode ocorrer na presença ou ausência de uma associação marginal entre a exposição e a desfecho. Além disso, o nome “collider bias” refere-se a qualquer forma de condicionamento por um colisor, incluindo pareamento, restrição, estratificação e ajuste em modelos de regressão.

Outros aspectos da amostragem e não-resposta podem ser, formalmente, tratados da mesma maneira (condicionamento por um colisor). A restrição a um único nível de um colisor S é análogo à restrição do conjunto de dados aos indivíduos com dados completos. Dados faltantes são comuns em estudos clínicos e

epidemiológicos, e podem surgir por uma variedade de razões. Podem resultar, por exemplo, da recusa de certos indivíduos de participarem da coleta de dados (“não-resposta da unidade”), ou por não fornecerem informações sobre uma medida específica (“não-resposta ao item”).

Como destacado em Sterne et al (2010), uma das principais dificuldades é determinar a abordagem mais adequada para tratar os dados faltantes. Vários métodos estatísticos têm sido propostos na literatura, incluindo a abordagem de exclusão de todos os indivíduos com dados faltantes (chamada de análise de caso completo), métodos de imputação única, como a última observação relatada (*Last Observation Carried Forward*), e abordagens mais avançadas, como as técnicas de imputação múltipla e formulações baseadas em verossimilhança (KARAHALIOS ET AL, 2012).

Cada uma das abordagens descritas anteriormente faz suposições sobre os dados em falta, que não podem ser verificados a partir dos dados observados. Sterne et al (2009) descrevem uma série de dificuldades que podem ser encontradas no procedimento de imputação, que podem levar a conclusões inadequadas para a análise epidemiológica de interesse. Deste modo, tem sido destacado na literatura a importância da representação das hipóteses sobre os dados faltantes por meio de diagramas causais e a utilização de ferramentas de análise de sensibilidade para avaliar a robustez dos resultados.

As análises estatísticas que excluem indivíduos com dados faltantes resultam em estimativas com menor precisão, em comparação com a análise de todos os indivíduos e, mais importante, podem implicar em viés. Isto porque os perfis de resposta dos não respondentes e dos entrevistados podem ser diferentes. Além disso, o efeito cumulativo de dados faltantes em algumas variáveis pode implicar a exclusão de uma proporção substancial da amostra original, o que, por sua vez, pode levar à uma perda substancial de precisão e poder (STERNE ET AL, 2010).

O risco de viés depende das razões pelas quais os dados estão faltando. As razões para a falta de dados são geralmente classificadas como: ausência totalmente aleatória (MCAR), ausência aleatória (MAR) e ausência não aleatória (MNAR).

No exemplo anterior (figura 7), podemos interpretar o vértice S como um indicador de dados faltantes para uma variável Z. Onde $S=1$ indicaria que os dados foram observados. Como descrito em Daniel et al (2012), a variável Z é MCAR se a

probabilidade de que Z seja observado, dado os dados completos V_0 , é independente de V_0 . Deste modo, a hipótese de ausência totalmente aleatória assume que $Pr(S=1 | V_0)$ é independente do valor de v_0 . Ou seja, que a probabilidade da falta de dados da variável Z não está relacionada com alguma outra variável medida e, é independente dos valores de Z . Em diagrama causal para esta situação, o vértice S (indicador de seleção/perdas) estaria isolado do resto do gráfico e, em geral, o condicionamento em $S=1$ não terá nenhuma consequência (em termos de viés de colisão).

Essas questões relacionadas ao confundimento, à mensuração de variáveis e à seleção da amostra tornam-se ainda mais complexas se considerarmos a dinâmica das relações causais no sistema de interesse. Há um crescente reconhecimento de que, para uma melhor compreensão e representação causal de muitos temas em epidemiologia, é necessário considerar a dependência entre os diferentes mecanismos que influenciam os desfechos em saúde e a dinâmica dessas relações ao longo do tempo (GALEA et al, 2010; AUCHINCLOSS et al 2008).

Obviamente, cada sistema de interesse pode ser representado considerando diferentes níveis de descrição. Assim, um modelo causal representa as hipóteses sobre os processos causais, em um nível particular de granularidade. Nesse contexto, um dos aspectos mais difíceis do processo de modelagem é a identificação do nível ideal de simplificação, que dependerá, não somente, das características do sistema em estudo, mas também dos objetivos da modelagem. Deste modo, considerando que a inferência causal tem como objetivo central prever os efeitos de intervenções em tais sistemas, então pode ser necessário avaliar o impacto dessas simplificações adotadas sobre a capacidade de se gerar conhecimento causal a partir desses modelos. Deste modo, é necessário considerar que, em algumas circunstâncias, a compreensão dos mecanismos envolvendo, por exemplo, ciclos causais e relações de causa e efeito que são distantes no tempo e no espaço pode ser essencial para a avaliação correta dos efeitos de certas intervenções.

Considere os estudos que avaliam os efeitos de variáveis relacionadas ao ambiente alimentar (por exemplo, a acessibilidade de alimentos - mensurada, pela proximidade a supermercados, restaurantes de *fast foods*, etc.) sobre o comportamento alimentar dos indivíduos. De acordo com a revisão sistemática mais recente sobre esse tema (CASPI et al, 2012), grande parte dos estudos apresentam

um desenho transversal. Deste modo, nessas investigações, os modelos causais assumem implicitamente que (i) o ambiente alimentar é relativamente constante ao longo da vida ou (ii) que o ambiente alimentar tem um efeito instantâneo sobre o desfecho.

No entanto, os indivíduos podem transitar entre diferentes ambientes ao longo do seu percurso de vida, e esses ambientes também podem mudar ao longo do tempo (MORLAND et al, 2014). Deste modo, se o acesso a supermercados e restaurantes muda ao longo do tempo, então uma única avaliação desses fatores ambientais pode não refletir o nível real de exposição individual. Além disso, é importante considerar o intervalo temporal entre essas exposições e seus possíveis efeitos sobre a dieta ou outros desfechos em saúde. Morland et al (2014) destaca que as mudanças no ambiente podem apresentar efeitos relativamente rápidos sobre o comportamento alimentar, enquanto os efeitos sobre outros desfechos, por exemplo, o IMC, hipertensão ou diabetes, necessitam de um tempo maior para se manifestarem. Deste modo, para cada sistema causal em estudo, deve-se avaliar o número de mensurações necessárias para caracterizar o nível de exposição ao longo do tempo e o intervalo necessário entre essas mensurações para que os efeitos causais possam ser observados.

Destaca-se que a indexação dos vértices no tempo está implícita em todos os DAG causais. Ou seja, assumimos que toda causalidade opera entre variáveis em pontos distintos no tempo. No entanto, a capacidade de decompor os ciclos causais em uma estrutura gráfica dinâmica depende, principalmente, da escala de tempo e da periodicidade da coleta de dados. Por exemplo, se os ciclos de retroalimentação ocorrem em períodos muito curtos, pode ser operacionalmente difícil representar essas relações em um DAG. Outras dificuldades conceituais e operacionais relacionadas aos ciclos causal e estruturas dinâmicas são discutidas em Dash (2003) e Aalen et al (2016).

Um possível problema em relação à classificação do tempo de exposição refere-se ao viés de tempo imortal. De forma geral, o termo “tempo imortal” é utilizado para se referir ao período de tempo de acompanhamento da coorte no qual o desfecho não pode ocorrer (ROTHMAN et al, 2008). O viés de tempo imortal pode surgir quando o período entre a entrada da coorte e a data da primeira exposição, durante o qual o evento de interesse não ocorreu, é classificado erroneamente ou simplesmente excluído e não contabilizado na análise.

Além disso, ao construir modelos causais é necessário que o pesquisador estipule o papel de cada variável nos processos que geram os dados. Deste modo, certas variáveis são definidas, por exemplo, como exposição, desfecho, confundidores ou mediadoras. A dimensão do tempo aumenta a complexidade dos modelos causais, pois não apenas os valores, mas também o papel dessas variáveis pode mudar ao longo do tempo. Nesse contexto, destacam-se, por exemplo, os confundidores tempo dependentes.

Por exemplo, nos estudos que avaliam o efeito do ambiente alimentar sobre o IMC, o comportamento alimentar é comumente considerado como uma variável mediadora (pois, o acesso a restaurantes de *fast foods* poderia incentivar o consumo excessivo de alimentos de alto densidade energética, enquanto a presença de supermercados levaria ao aumento da ingestão de frutas e vegetais – e ambos fatores poderiam influenciar o IMC). No entanto, o comportamento alimentar pode ter um efeito sobre futuras localizações residenciais e, por consequência, pode influenciar a exposição a esses fatores ambientais (se as pessoas com comportamentos alimentares mais saudáveis optam por viver em áreas com melhor acesso aos alimentos) (MORLAND et al, 2014). Neste exemplo, o comportamento alimentar desempenha simultaneamente o papel de variável mediadora e de confusão. Nesses casos, como demonstrado em Robins (1986), a estimação de efeitos causais por meio de técnicas estatísticas tradicionais pode levar a conclusões inadequadas, pois o ajuste por essas variáveis (confundidores tempos dependentes) é necessário para o controle de confundimento, mas o condicionamento por essas variáveis, por meio de métodos baseados em estratificação, podem implicar na introdução de viés de colisão. Nesses casos, é indicada a utilização de ferramentas alternativas como os métodos de *g-estimation* e técnicas baseadas na ponderação pela probabilidade inversa (IPW).

Outro desafio para inferência causal refere-se à possibilidade de interferência (ou contágio) entre os indivíduos na população de estudo. De forma geral, a interferência pode ser definida como qualquer processo pelo qual o status de exposição de um indivíduo pode afetar os desfechos dos demais indivíduos na população. Nesse contexto, o efeito causal para um indivíduo será uma mistura entre o efeito da exposição de interesse e o efeito devido a interação com as suas unidades “vizinhas”. Grande parte dos modelos formais para causalidade e das técnicas estatísticas tradicionais adotam a hipótese de não interferência entre

unidades (OGBURN et al, 2014). No entanto, como destacado em Ronir & Struchiner (1995), “a ocorrência de doenças contagiosas constitui um bom exemplo em que essa hipótese é claramente violada, já que estas doenças têm como característica principal a presença de cadeias de transmissão de um hospedeiro para outro, de modo que um indivíduo (unidade) eventualmente torna-se infectado dependendo de contatos com outros já infectados”. Contudo, a interferência entre indivíduos não se restringe ao contexto das chamadas “doenças transmissíveis”. A interferência entre unidades pode ocorrer, por exemplo, em ambientes em que os desfechos dos indivíduos da população acontecem por meio de interações sociais.

Há diferentes mecanismos relacionados ao fenômeno de interferência que podem implicar em diferentes estratégias de identificação e de estimação dos efeitos causais (efeito total e efeito entre pares)³. Em alguns cenários a interferência pode ocorrer apenas de forma parcial, isto é, dentro de subgrupos de indivíduos (por exemplo, quando os subgrupos são separados no espaço ou no tempo e a influência entre unidades ocorre apenas entre os indivíduos do mesmo grupo). Mesmo nesse cenário simples, a presença de interferência pode ter implicações para o controle de confundimento, para a mensuração de variáveis e para seleção da amostra. No entanto, se as relações sociais entre os indivíduos em uma população são mais complexas a hipótese de interferência parcial será menos plausível. Por exemplo, em situações onde os indivíduos estão relacionados entre si através de uma série de laços que podem ser mapeados por uma rede social (OGBURN et al, 2014). Nesse contexto, a configuração de uma rede social também oferece certos desafios conceituais e metodológicos.

3 Há diferentes terminologias para os efeitos que resultam da interferência entre os indivíduos na população de estudo. No contexto das doenças infecciosas é comumente utilizado o termo “efeitos indiretos”. Por exemplo, em Halloran e Struchiner (1991), o “efeito indireto da vacinação” é utilizado para descrever os processos pelos quais a vacinação de uma pessoa pode afetar o desfecho (a infecção) de outro indivíduo na população. Na literatura da inferência causal, os efeitos devido à interferência também são chamados de “efeitos colaterais” (*Spillover effects*). Na literatura das ciências sociais, o fenômeno da interferência é comumente referido como “interação social” – por exemplo, em estudos envolvendo grupos (bairros, salas de aulas) nos quais os indivíduos interagem uns com os outros de tal maneira que o desfecho de uma pessoa depende da exposição das demais. Os efeitos que resultam dessas interações são descritos como efeitos de pares (*peer effects*), efeitos de grupos (*group effects*), efeitos relacionais (*relational effects*) ou efeitos de rede social (*network effects*).

3 COMUNICAÇÃO DE ESTUDOS EPIDEMIOLÓGICOS

Há diferentes maneiras de se avaliar um estudo epidemiológico. Uma delas refere-se à qualidade metodológica do estudo; ou seja, se os aspectos relacionados à coleta e análise de dados conferem um alto grau de garantia às inferências de um estudo⁴. Existem diferentes ferramentas para a avaliação da qualidade metodológica (ou risco de viés) de estudos observacionais, que mostram uma grande variabilidade no processo de desenvolvimento, no número de itens incluídos, tempo e esforço exigido e características de conteúdo (DEEKS et al, 2003; SANDERSON et al, 2007; GROENWOLD & ROVERS, 2010; LANG & KLEIJNEN, 2010). Em todos os casos, avaliar a qualidade de um estudo epidemiológico é um processo desafiador que requer um rigoroso conhecimento metodológico e clínico (DOI & WILLIAMS, 2013)

Outro aspecto refere-se à qualidade da comunicação⁵; isto é, a integralidade com que um estudo é apresentado e se itens importantes para a avaliação adequada da validade interna e externa são claramente relatados. O relato transparente é fundamental para a avaliação da qualidade, confiabilidade e relevância da pesquisa por parte da comunidade científica e pelo público geral. Deste modo, sem um relato científico adequado, transparente e completo, os recursos de pesquisa são potencialmente desperdiçados, o valor da pesquisa é diminuído e o processo científico não cumpre as suas obrigações éticas (DOI & WILLIAMS, 2013).

A necessidade de se estabelecer padrões mínimos de qualidade do relato, resultou no desenvolvimento de diretrizes (como a declaração STROBE – *Strengthening the Reporting of Observational Studies in Epidemiology*, von Elm et al,

4 Como descrito em Doi & Williams (2013), existem várias definições do conceito de “qualidade do estudo”. No contexto das revisões sistemáticas, a avaliação da qualidade de estudos primários refere-se, principalmente, à validade interna do estudo e é descrito mais precisamente como avaliação do risco de viés. O termo “Qualidade metodológica” também é utilizado neste contexto.

5 Como destacado em Gough et al (2012), há diferentes dimensões de qualidade nas quais a pesquisa pode ser julgada, dependendo da perspectiva e das perguntas dos avaliadores. Por exemplo, se considerarmos como os resultados da pesquisa podem ser utilizados ou aplicados, outras dimensões da qualidade tornam-se pertinentes, em particular, a relevância dos achados. Além de produzir resultados que são úteis em um contexto aplicado, a realização de um estudo também pode levar a outros benefícios, como novas habilidades e aprendizagem para a equipe envolvida.

2007) que especificam um conjunto mínimo de itens para uma comunicação completa e transparente de estudos transversais, de caso controle e de coorte, agregando as semelhanças entre esses delineamentos e especificando por tipo de estudo, quando necessário (GALVÃO et al, 2016). A declaração STROBE foi publicada em conjunto com um documento que apresenta uma explicação detalhada de cada item da lista de verificação (VANDENBROUCKE et al, 2009).

Como destacado em Malta et al (2010), a Iniciativa STROBE tem suscitado um crescente debate na literatura científica. Para alguns autores, a iniciativa é percebida como uma estratégia importante que pode contribuir para a melhoria na qualidade do relato de estudos observacionais (EBRAHIM & CLARKE, 2007). No entanto, alguns editores se mostram reticentes diante do fato da Iniciativa buscar formalizar o relato em um campo tão heterogêneo como a epidemiologia observacional (*Probing STROBE*, 2007 – Editorial do *Epidemiology*). Uma das preocupações é que este guia pode evoluir de um conjunto de diretrizes de relato para uma ferramenta de avaliação da qualidade metodológica e, que esta distinção (entre qualidade do relato e qualidade do estudo) não é descrita de forma clara nos documentos do STROBE.

Da Costa et al (2011) examinaram o contexto e as circunstâncias em que a declaração STROBE foi utilizada em uma amostra de 100 artigos que citavam a declaração. Entre os 100 artigos incluídos, apenas 32 eram de estudos observacionais; dentre esses, 26 fizeram uso adequado do STROBE (como uma diretriz para o relato), e três usos foram considerados inapropriados. Entre as 19 revisões sistemáticas avaliadas, 10 utilizaram a diretriz como uma ferramenta para avaliar a qualidade metodológica de estudos observacionais.

Além disso, Rothman & Poole (2007) destacam que a Iniciativa STROBE deve ser vista como um processo de constante revisão, uma vez que qualquer estratégia desse tipo representa um consenso de um grupo específico, em um determinado momento.

De acordo com o site da Iniciativa STROBE (até novembro de 2016), 122 revistas referem-se à Declaração STROBE na seção de Instruções para os Autores. Além disso, há diferentes adaptações ao STROBE disponíveis na literatura. A primeira extensão foi desenvolvida para estudos de Associação Genética (STREGA) (LITTLE et al, 2009). Mais recentemente, foram publicadas outras extensões de diretrizes para estudos de Epidemiologia Molecular (STROBE-ME – GALLO, et al

2012), Epidemiologia Molecular para doenças infecciosas (STROME-ID – FIELD et al, 2014), Epidemiologia Nutricional (STROBE-nut – LACHAT et al, 2016), Pesquisas em Infecção neonatal (STROBE-NI – FITCHETT et al, 2016) e Epidemiologia Veterinária (STROBE-Vet – SARGEANT et al, 2016), dentre outras.

3.1 Comunicação de fontes de incerteza

O tema da (in)certeza tem sido investigado sob diferentes perspectivas ao longo da história. Discussões comuns sobre o tema da incerteza podem envolver a caracterização de sua natureza ou de suas fontes, os aspectos relacionados a percepção e ao raciocínio sob incerteza (p.ex., KAHNEMAN & TVERSKY, 1992; ARKES & HAMMOND, 1986), as formas de avaliação ou quantificação (p.ex., LASH, et al, 2007) e as maneiras pelas quais as pessoas expressam a incerteza (p.ex., BITTI et al, 2014).

Há na literatura diferentes classificações de incertezas quanto a sua natureza ou fonte (KAHNEMAN & TVERSKY, 1982; ÜLKÜMEN et al., 2016; SCHNEIDER, et al., 1998). Para alguns autores, a incerteza pode ser classificada em duas categorias básicas: a incerteza epistêmica e a incerteza linguística (REGAN et al, 2002). A incerteza epistêmica estaria associada ao conhecimento do estado de um sistema e resulta das limitações do conhecimento ou dos métodos. A incerteza linguística, por outro lado, surge porque grande parte da nossa linguagem apresenta usos indefinidos e alternativos. Deste modo, as regras da linguagem podem ser ambíguas, vagas e contexto-dependentes. Outras discussões teóricas sobre as relações entre imprecisão, incerteza e certeza são apresentadas em Debrock & Hulswit (2013) e Wittgenstein (1969).

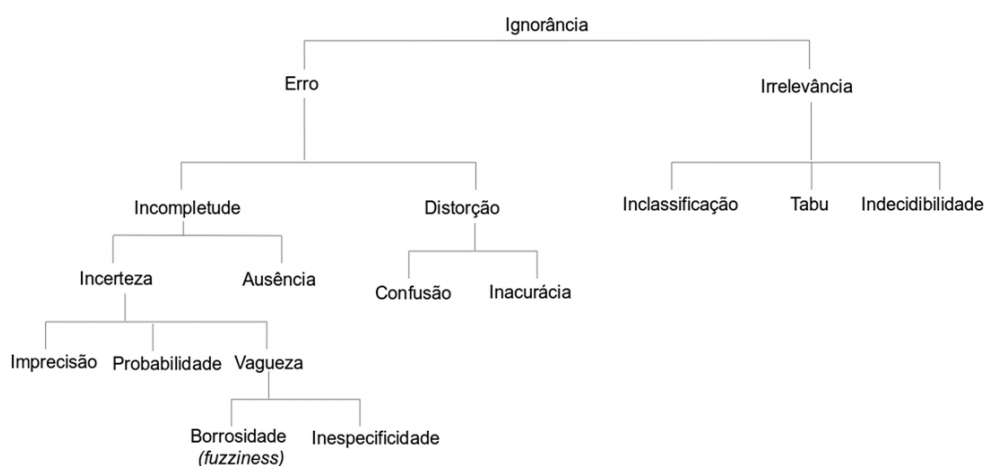
Na área relacionada ao raciocínio sob incerteza, outra classificação comum consiste na distinção entre a incerteza interna e externa (KAHNEMAN & TVERSKY, 1982A; FOX & ÜLKÜMEN, 2011). A incerteza interna decorre da falta de conhecimento pessoal e a incerteza externa é atribuível aos processos do mundo externo (KAHNEMAN & TVERSKY, 1982). Como destacado em Kahneman & Tversky (1982a) a distinção entre incerteza interna (ignorância) e externa está intimamente relacionada à distinção mais geral entre atribuições internas e externas

da experiência. Por exemplo, a incerteza sobre eventos passados é geralmente experimentada como ignorância (incerteza interna), especialmente se os fatos são apenas conhecidos por outra pessoa, enquanto as incertezas sobre o futuro são mais frequentemente associadas às disposições do sistema (incerteza externa) (KAHNEMAN & TVERSKY, 1982a)

Outros autores descrevem duas variantes de incerteza subjetiva: epistêmica e aleatória (ÜLKÜMEN et al., 2016; HACKING, 1975). Por exemplo, para Ülkümen et al. (2016) a incerteza epistêmica é caracterizada pelo conhecimento, habilidades ou métodos inadequados, enquanto a incerteza aleatória é caracterizada por comportamentos estocásticos. A aleatoriedade inerente em um sistema ocorre, não por conta da nossa compreensão limitada dos processos, mas porque esses processos são, em princípio, irreduzíveis a um sistema determinista. Uma questão recorrente sobre a separação entre variabilidade aleatória e incerteza epistêmica é o debate sobre se algumas incertezas podem ser eliminadas, dado o tempo e esforços suficientes, enquanto outras podem nunca serem resolvidas.

Outra abordagem para estruturar a incerteza foi apresentada em Smithson (1989) na forma de uma taxonomia da ignorância (Figura 8).

Figura 8. Taxonomia de ignorância de Smithson.



Fonte: SMITHSON, 1989,

Smithson distingue dois tipos fundamentais de ignorância: o estado de ignorância (erro) e o ato de ignorar (irrelevância). O erro pode surgir de informações incompletas ou distorcidas (ou ambos) e a irrelevância refere-se às questões que

são deliberadamente ou inconscientemente ignoradas. Como destacado em (SMITHSON,1989), estes dois tipos de ignorância denotam estratégias diferentes de lidar com anomalias e outras ameaças à ordem cognitiva estabelecida, ou seja, a inclusão e a exclusão. Uma estratégia é ignorá-los ou não os perceber, com isso declarando as anomalias irrelevantes e banindo-as da realidade. A outra estratégia é rever a nossa estrutura da realidade para dar lugar ao material anômalo, mesmo que seja apenas por meio de uma declaração de ignorância.

Smithson (1989) divide a irrelevância em três subcategorias, a “não classificação”, o tabu e a indecidibilidade. A não classificação refere-se às intuições que as pessoas carregam com elas e negociam com outras pessoas sobre como seus domínios cognitivos se encaixam. A indecidibilidade é atribuída a questões que as pessoas são incapazes de designar verdadeiras ou falsas, quer porque consideram o problema insolúvel, ou porque a questão da validade não é pertinente. O tabu, por outro lado, é uma irrelevância socialmente imposta.

Com relação à “incompletude”, Smithson (1989) diferencia entre a incompletude em grau, que ele chama de “incerteza”, e a incompletude em espécie, que ele chama de “ausência”. Por sua vez, a incerteza pode ser subdividida em “imprecisão”, “probabilidade” e “vagueza”. Para Bammer & Smithson (2008), “a imprecisão relaciona-se a uma gama de valores possíveis num continuum, a probabilidade refere-se às leis do acaso e a ambiguidade refere-se a um número finito de possibilidades distintas”.

Existem diferentes adaptações e alternativas às classificações apresentadas anteriormente. No entanto, o ponto a ser destacado aqui é que uma consideração mais profunda sobre esse tópico sugere que a incerteza é muito mais do que uma questão estatística. Como destacado em Briggs et al (2009), ela também se relaciona com a maneira como percebemos e interagimos com o mundo. Além disso, contrário à visão da ignorância como primariamente negativas, o engajamento humano com a incerteza é quase sempre um empreendimento motriz misto (BAMMER & SMITHSON, 2008). Além disso, também representa um produto da ciência, no sentido de que a pesquisa científica pode ser vista como uma atividade de refinamento da ignorância (FIRESTEIN, 2012).

No entanto, diferentes disciplinas estabelecem diversas formas de lidar com a incerteza. Uma possível diferença significativa entre áreas, destacada em Bammer & Smithson (2008), diz respeito à urgência associada às decisões; algumas incertezas,

como aquelas relacionadas à investigação de um surto de doença, requerem avaliações e respostas rápidas, enquanto outras incertezas, relacionadas à investigação de um incidente histórico ou medições físicas, podem ser abordadas de uma forma mais lenta.

Na epidemiologia tem sido destacada a necessidade de compreender e relatar a incerteza de forma muito mais sistemática. Como destacado em Lash et al (2011), se o objetivo da investigação epidemiológica é obter uma estimativa válida, precisa e generalizável do efeito de uma exposição sobre a ocorrência de um desfecho, então os investigadores têm uma dupla obrigação: primeiro, eles devem projetar suas investigações visando melhorar a precisão e validade da estimativa de efeito de interesse. Em segundo lugar, reconhecendo que nenhum estudo é perfeito, eles devem informar as partes interessadas o quão perto eles acreditam que as estimativas possam estar desses objetivos de precisão e validade.

A comunicação transparente das fontes de incerteza de um estudo é fundamental para interpretação da validade dos resultados por parte da comunidade científica e pelo público geral e pode ter um papel fundamental no direcionamento de pesquisas futuras. No entanto, pesquisas anteriores sugerem que poucos trabalhos na literatura médica e epidemiológica discutem abertamente as limitações do conhecimento ou dos métodos (IOANNIDIS, 2007; POCOCK et al, 2004; GROENWOLD et al, 2008; PAPATHANASIOU et al, 2010; SHARP et al, 2014; KARAHALIOS et al, 2012; POUWELS et al, 2015).

Por exemplo, com relação a comunicação de confundimento, Groenwold et al (2008) avaliaram artigos publicados em cinco revistas médicas gerais e cinco revistas epidemiológicas no período entre 2004 a 2007. Dos 174 artigos incluídos na análise, os detalhes sobre o processo de seleção dos fatores de confusão foram relatados em apenas 10% dos estudos. Além disso, 60% dos estudos relataram a possibilidade de confundimento residual, mas apenas 9% comentaram sobre o seu potencial impacto sobre as estimativas apresentadas. Resultados semelhantes foram observados em Pouwels et al (2015) ao comparar a qualidade da comunicação sobre confundimento em artigos publicados antes e após a declaração STROBE.

Ao avaliar a comunicação sobre dados faltantes em estudos longitudinais, Karahalios et al (2012) observaram que entre os artigos incluídos na revisão (n=82), apenas 43% (n=35) relataram a quantidade de dados faltantes (que variou de 4 a

80% com uma média de 25%), e 66% (n=54) dos estudos realizaram análises de casos completos. Resultados semelhantes foram observados com relação à comunicação de erros de mensuração. Jurek et al. (2006) avaliaram uma amostra (n=57) de artigos publicados em três revistas epidemiológicas e observaram que 61% (n=35) relataram a possibilidade de erros de mensuração da exposição, mas apenas 28% (n=16) avaliaram qualitativamente o possível impacto desses erros nas estimativas de efeito.

Uma prática comum na epidemiologia consiste na apresentação dos resultados em três partes: uma estimativa pontual (p.ex. o risco relativo), uma avaliação estatística da incerteza (p.ex. o intervalo de confiança) e uma descrição qualitativa sobre as fontes de erros sistemáticos que podem ter afetado a validade do estudo (LASH, 2007). Essas avaliações qualitativas são normalmente apresentadas como “limitações do estudo” na seção de discussão do artigo científico.

Em um estudo que avaliou o efeito da revisão por pares sobre a qualidade final do manuscrito, Goodman & Riet (2009) relataram que entre os 34 itens avaliados, o reconhecimento das limitações do estudo foi o item mais problemático na apresentação dos manuscritos. Por meio de pesquisa automática de palavras em uma amostra de 400 artigos publicados em seis revistas científicas de alto impacto (IOANNIDIS, 2007), foi observado que apenas 17% das publicações contém termos relacionados à limitação do estudo. Em uma avaliação mais detalhada da literatura, em que dois revisores avaliaram as seções de resumo e de discussão de 300 artigos publicados em revistas biomédicas, Ter Rie et al, 2013 observaram que 81 das 300 publicações não reconheceram qualquer limitação do estudo (27%) e 219 publicações reconheceram uma mediana de 3 limitações (1-8) na seção de discussão (73%).

Na literatura epidemiológica também têm sido destacadas algumas limitações da abordagem tradicional de avaliação qualitativa das fontes de erros sistemático (LASH, 2007; GREENLAND, 2005; DIEZ-ROUX, 2011). Uma delas é que as avaliações qualitativas dependem, muitas vezes, de heurísticas simplificadoras, que podem resultar em erros de avaliação (GREENLAND, 2005), p.ex., a de que todos os erros de mensuração não diferenciais influenciam os valores esperados de um estimador em direção ao valor nulo.

Como destacado em Lash et al (2011), a avaliação da magnitude dos erros sistemáticos, em comparação com a magnitude dos efeitos da exposição, torna-se, portanto, um exercício de raciocínio sob incerteza; e a capacidade humana de raciocinar sob incerteza é suscetível a erros sistemáticos de julgamento. Deste modo, de acordo com os autores “a abordagem qualitativa frequentemente não protegerá contra as tendências em favorecer os efeitos da exposição sobre os erros sistemáticos, como uma explicação plausível para as associações observadas”.

Além disso, dado que algumas fontes de vieses nem sempre são incorporadas na modelagem, as medidas de incerteza comumente apresentadas não refletem todas as fontes de incerteza sobre o modelo. Como destacado em Greenland (2005), a omissão inevitável de alguma fonte de viés não pode justificar a omissão de todos os tipos possíveis vieses. Diferentes estratégias metodológicas têm sido apresentadas para análise quantitativa de viés em dados epidemiológicos (LASH ET AL, 2011; GREENLAND, 2005). Avaliações quantitativas de viés podem fornecer informações valiosas sobre a importância de várias fontes de erro e podem contribuir para uma comunicação mais transparente da incerteza sobre resultados do estudo.

Deste modo, podemos destacar, pelo menos, 4 abordagens para a comunicação de fontes de viés em estudos epidemiológicos:

1. Ignorar a possibilidade de erros sistemáticos (ou assumir o cenário de melhor caso em que os erros são mínimos e podem ser desconsiderados);
2. Reconhecer algum tipo de viés em potencial;
3. Avaliar qualitativamente o impacto (i.e, magnitude e direção) do viés em potencial nos resultados do estudo;
4. Avaliar quantitativamente o impacto do viés nos resultados do estudo.

3.1.1 Marcadores linguísticos de incerteza

Como destacado em Briggs et al (2009), as narrativas também são meios poderosos para transmitir e discutir incertezas. O nível de incerteza pode ser comunicado, não apenas, pela presença de certas informações no texto, mas também pela maneira na qual os autores expressam essas informações.

Existem vários fenômenos linguísticos que são referidos como incerteza na literatura (FRASER, 2010; NUYTS, 2001; CLEMEN, 1997). No entanto, ao longo dos últimos anos, o uso de *hedging* tem atraído a atenção crescente de pesquisadores que investigam a escrita acadêmica (GILLAERTS & VAN DE VELDE, 2010; HYLAND & TSE, 2004; SALAGER-MEYER, 1997).

Como sugerido em Fraser (2010), o conceito de hedging foi apresentado pela primeira vez no contexto linguístico em Weinreich (1966) e popularizado pelos trabalhos de George Lakoff (1972). Para Lakoff (1973) a função dos termos *hedges* “é tornar as coisas mais ou menos difusas”. A investigação de Lakoff foi baseada na teoria dos conjuntos difusos (*Fuzzy sets*) desenvolvida em 1964 por Lotfi A. Zadeh, para tratar de problemas de classificações de categorias que não possuem fronteiras de pertinência bem definidas.

De forma geral, a noção de *hedging* pode ser vista como qualquer recurso linguístico utilizado para indicar: (a) a falta de um comprometimento completo com o valor de verdade de uma sentença acompanhante ou (b) um desejo de não expressar esse comprometimento categoricamente (HYLAND, 1998). Deste modo, os termos *hedges* são utilizados principalmente na aproximação e atenuação de expressões e são um meio pelo qual os autores podem expressar a extensão da incerteza sobre a importância e validade dos achados ou hipóteses.

No entanto, alguns autores apresentaram críticas em relação ao uso intensivo de *hedging* (PINKER, 2014; ROLAND, 2007). Por exemplo, Roland (2007) argumenta que o uso de *hedges* impede que os pesquisadores se comprometam com o assunto da pesquisa, deixando a interpretação aos jornalistas ou qualquer outra parte interessada e, portanto, criam as condições para manipulação e interpretação errada das informações. No entanto, Salager-Meyer (1994) argumenta que o uso de *hedging* não deve ser considerado como uma tática de encobrimento, pois pode refletir um esforço genuíno do autor de expressar a incerteza científica. Os autores podem querer reduzir a força de suas reivindicações simplesmente por acreditarem que essas reivindicações mais fortes não seriam justificadas (SALAGER-MEYER, 1997). Salager-Meyer (1997) avaliou a distribuição de *hedges*

em diferentes seções dos artigos de pesquisa médica, indicando uma maior atenuação da linguagem nas seções de discussão dos artigos, e menor frequência de hedges na seção de métodos. Além disso, alguns resultados indicam que os *hedges* ocorrem com mais frequência em contextos de citação (DI MARCO et al, 2006).

Na sua forma mais simples, os *hedges* são palavras isoladas, como advérbios (geralmente, provavelmente), verbos (sugere, indica) e verbos modais (podem, devem). A Tabela 2 apresenta alguns exemplos comuns de termos *hedges* apresentados em Hyland (1998; 2004).

Tabela 2. Exemplos de *hedges* utilizados em artigos científicos.

Tipo	Exemplos
Auxiliares modais	<i>May, might, could, would, should</i>
Verbos epistêmicos de julgamento	<i>Suggest, indicate, speculate, believe, assume</i>
Verbos epistêmicos evidenciais	<i>Appear, seem</i>
Verbos epistêmicos dedutivos	<i>Conclude, infer, deduce</i>
Adjetivos epistêmicos	<i>Likely, probable, possible</i>
Advérbios epistêmicos	<i>Probably, possibly, perhaps, generally</i>
Substantivos epistêmicos	<i>Possibility, suggestion</i>

Fonte: HYLAND, 1998; 2004.

No entanto, em muitos casos, as palavras identificadas por um conjunto de pistas lexicais (tabela 2) podem não apresentar a função de atenuação de uma sentença acompanhante. Por exemplo, na avaliação realizada em Prokofieva & Hirschberg (2014), a pesquisa automática baseada em pistas lexicais apresentou uma precisão de apenas 45% em comparação ao padrão ouro (corpus anotado).

Além disso, como sugerido em Clemen (1997), dependendo do contexto, grande parte dos itens ou expressões linguísticas podem adquirir a qualidade de *hedges*. Nos últimos anos, o conjunto de itens considerados como *hedges* tem sido expandido para abranger uma ampla gama de fenômenos, e não há uniformidade entre os estudos sobre quais fenômenos linguísticos devem ser considerados como

hedges (VARTTALA, 2001; FRASER, 2010). Embora as listas apresentadas em estudos anteriores possam não indicar todos os potenciais casos de *hedging*, elas são úteis na seleção de alguns dos fenômenos linguísticos centrais à essa estratégia (VARTTALA, 2001).

3.2 Possíveis diferenças na percepção e no relato de incertezas

Parece razoável supor que as atitudes em relação à comunicação de incertezas em artigos científicos dependem de fatores ligados à sua percepção e aos incentivos e restrições à sua comunicação. Diferenças na percepção de incerteza podem estar associadas a características individuais, como complexidade cognitiva (BIERI, 1955), vieses cognitivos (TVERSKY & KAHNEMAN, 1974), traços de personalidades (HODSON, 1999), bem como ao conhecimento específico do domínio e crenças sobre o conhecimento (BUEHL, 2001).

Duncan (1972) sugere que indivíduos com tolerância à incerteza podem perceber situações como menos incertas do que os indivíduos com baixa tolerância. Os aspectos relacionados a múltiplas dimensões da percepção individual e dos processos de avaliação têm sido rotulados como estudos de complexidade cognitiva (BIERI, 1955; DOWEY et al, 1977; 1972). Outros autores sugerem possíveis relações entre a tolerância a incerteza e o processo criativo (Kornilova et al 2010; KITAEV-SMYK, 2007). Outras relações entre aspectos da cognição e traços de personalidade também têm sido sugeridas. Por exemplo, o neuroticismo parece estar associado a um desempenho inferior em vários domínios, incluindo o processamento de informações, análise de padrões e memória que, por sua vez, podem influenciar a maneira de lidar com a incerteza (HODSON, 1999).

A pesquisa psicológica revelou que os indivíduos buscam frequentemente a manutenção da consistência interna, o que pode levar a mecanismos cognitivos, como a dissonância cognitiva e o viés de confirmação (NICKERSON, 1998; FESTINGER, 1962). No contexto acadêmico, os conflitos de interesses mais discutidos são aqueles decorres de incentivos financeiros (BEKELMAN et al, 2003). No entanto, variações do viés de confirmação podem ocorrer por outros fatores não diretamente relacionados a questões financeiras, como opiniões políticas e crenças

personais (MONTIBELLER & WINTERFELDT, 2018). Por exemplo, Cope & Allison (2010) sugerem o termo *white hat bias* para descrever o viés resultante da distorção de informações a serviço do que pode ser percebido como um fim positivo ou “politicamente correto”. Um dos fatores destacados inclui a citação seletiva de resultados que favorecem apenas um determinado ponto de vista.

Alguns autores distinguem dois tipos de vieses individuais: o viés cognitivo e o viés motivacional (MONTIBELLER & WINTERFELDT, 2018). De forma geral, um viés cognitivo é uma discrepância sistemática entre a resposta “correta” (dada por uma regra formal) em uma tarefa de julgamento e a resposta do tomador de decisão. Os vieses motivacionais podem ser definidos como aqueles em que os julgamentos são influenciados pela desejabilidade ou indesejabilidade de certos eventos, resultados ou escolhas (MONTIBELLER & WINTERFELDT, 2015; KUNDA, 1990). Deste modo, as atitudes em relação à comunicação de incertezas também podem ser influenciadas pela percepção dos riscos relacionados à sua comunicação. Por exemplo, Puhon et al (2012) argumentam que na atual cultura de publicação de estudos biomédicos, muitos autores podem não querer discutir as limitações de seus estudos por perceberem um limiar de transparência para além do qual a probabilidade de aceitação do manuscrito diminui.

Montibeller & Winterfeldt (2015) fornecem uma revisão de alguns vieses cognitivos e motivacionais na análise de risco e estratégias de mitigação. No contexto do viés de confirmação, algumas estratégias de mitigação sugeridas incluem o uso de múltiplos especialistas com diferentes pontos de vista, a utilização de análises contrafactuais nas avaliações de probabilidade, e a busca por evidências para as hipóteses alternativas (MONTIBELLER & WINTERFELDT, 2015; 2018). No entanto, a mitigação de alguns vieses pode exigir mudanças mais estruturais no sistema de ensino e a adoção de incentivos que reforçam avaliações mais completas e transparentes das informações (CROSKERRY et al, 2013a; 2013b).

Um passo considerado crítico na mitigação de vieses motivacionais e cognitivos é a consciência da existência do viés e de suas causas (CROSKERRY et al, 2013a), o que requer habilidades metacognitivas (isto é, a capacidade de refletir sobre os nossos próprios pensamentos e comportamentos). No entanto, o acesso consciente a esses processos parece ser limitado (FRITH, 2012). Por exemplo, Kruger e Dunning (1999) sugeriram que, em muitos domínios intelectuais e sociais, os indivíduos de pior desempenho possuem autoavaliações mais precisas e

superestimadas de suas habilidades e desempenhos. Esse viés de meta ignorância ficou conhecido como efeito Dunning-Kruger. Como sugerido em Dunning (2011), a meta-ignorância (ou ignorância da ignorância) surge porque a falta de conhecimento muitas vezes se esconde no reino dos “unknowns-unknowns”. Além disso, os autores sugerem que a especialização necessária para julgar o desempenho em muitos setores intelectuais e sociais, são exatamente os mesmos conhecimentos necessários para produzir um bom desempenho em primeiro lugar. No entanto, outros autores apresentam explicações alternativas ao possível fenômeno Dunning-Kruger, incluindo aspectos como regressão à média, erros de mensuração e outras críticas metodológicas (p.ex, KRUEGER & MUELLER, 2002). Outras considerações sobre possíveis interdependências entre a metacognição e o desempenho são apresentadas em Fleming & Dolan (2012) e Frith (2012).

Outras características individuais e aspectos contextuais, como restrições de tempo e o estresse, podem estar relacionados à percepção de incertezas. Por exemplo, estudos sobre o processo de geração de hipóteses sugerem que os indivíduos tendem a considerar apenas um subconjunto do total de hipóteses alternativas possíveis, e aquelas que são geradas tendem a ter a maior probabilidade *a priori* (THOMAS et al, 2008). Além disso, a quantidade de tempo disponível para a geração de hipóteses parece determinar, parcialmente, quantas hipóteses são geradas e quando o processo de geração de hipóteses é encerrado (Thomas et al, 2008). Além disso, possíveis fontes de diferenças individuais no pensamento crítico-analítico também têm sido sugeridas, incluindo fatores genéticos, a exposição a múltiplos idiomas e o estresse (BOLGER et al, 2014).

Além de vieses individuais, os vieses sociais ou de grupo podem afetar as atitudes e expectativas sobre como o trabalho científico deve ser conduzido e comunicado. Como sugerido em Veldkamp et al (2017), uma forte crença na imagem idealizada do cientista e a tendência de atribuir níveis mais altos de características desejáveis a pessoas do próprio grupo pode influenciar a adoção de práticas para redução de vieses em pesquisa.

Veldkamp et al (2017) analisaram a atribuição (intra e entre grupos) de traços relativos aos ideais científicos em uma amostra composta por cientistas de diferentes países (N=1088) e não cientistas americanos com alto nível de escolaridade (n=355). No grupo de cientistas foram observadas maiores atribuições (aos cientistas) de traços de objetividade, inteligência e comunalidade, e menores

diferenças nas atribuições de racionalidade e integridade. Além disso, os autores sugerem um possível viés de grupo relativo ao nível de carreira e ao gênero do pesquisador: cientistas estabelecidos atribuíram níveis mais elevados de traços científicos para cientistas do mesmo grupo do que para cientistas em início de carreira e estudantes de doutorado, e níveis mais altos para estudantes de doutorado do que para cientistas em início de carreira (VELDKAMP et al, 2017). Além disso, cientistas do sexo feminino atribuíram maiores níveis de traços científicos para cientistas do sexo feminino (VELDKAMP et al, 2017).

O conceito de pensamento de grupo (*groupthink*) foi sugerido em Janis (1972) para descrever certos fenômenos psicológicos que podem ocorrer na tomada de decisão em grupo quando certas condições antecedentes estão presentes, como a alta coesão do grupo, ameaças situacionais, homogeneidade de valores, isolamento dos grupos, dentre outros. O modelo apresentado em Janis (1972) consiste em quatro componentes: as condições antecedentes do pensamento de grupo (“causas potenciais”), um conjunto de sintomas do pensamento de grupo e sintomas de vieses na tomada de decisão (“mediadores”) e possíveis desfechos relativos à qualidade da decisão. Uma das hipóteses é que essas condições antecedentes podem provocar uma tendência à conformidade que, por sua vez, pode manifestar-se em um conjunto de sintomas de pensamento de grupo, incluindo: a super confiança ou super estimação das características e valores intra grupo (p.ex., como mais racionais ou moralmente superiores), estereótipos negativos correspondentes em relação ao grupo externo (como menos racionais ou moralmente inferiores), mente fechada (p.ex., racionalização da dúvida) e pressões por uniformidade (p.ex., pressão direta sobre os discordantes, auto censura, ilusão de consenso) (PARK, 1990). Os possíveis efeitos dos sintomas do pensamento de grupo incluem o levantamento incompleto de alternativas e objetivos, falhas na avaliação dos riscos de determinada escolha e a seletividade na busca de informação (PARK, 1990; JANIS, 1972). No contexto da modelagem, o pensamento de grupo pode refletir, por exemplo, em uniformidade dos métodos e técnicas de modelagem utilizadas em determinadas áreas de pesquisa (HAMALAINEN, 2015). No entanto, as evidências empíricas para algumas hipóteses do modelo de pensamento de grupo têm sido inconsistentes; além disso, críticas metodológicas e modelos alternativos também têm sido apresentados na literatura (ROSE et al, 2011).

3.2.1 Revisão editorial e diretrizes de relato

No contexto geral da qualidade dos relatórios de pesquisa, os fatores mais frequentemente discutidos na literatura incluem as características do processo de revisão por pares e aspectos da política editorial (JEFFERSON et al, 2002; STEVENS et al, 2014; KOVANIS et al, 2017; BRUCE et al, 2016; PASTORINO et al, 2016), o uso de diretrizes de relato (FULLER et al, 2015; BRUCE et al, 2016), além de outros aspectos mais estruturais relacionados ao sistema de ensino e de recompensa (NOSEK et al., 2012; IOANNIDIS, 2014; MARUSIĆ & MARUSIĆ, 2003).

Em uma revisão sistemática de ensaios clínicos randomizados (n=22) que avaliaram intervenções no processo de revisão por pares, Bruce et al (2016) destacam que há poucas evidências empíricas disponíveis sobre o papel da revisão editorial como um mecanismo para garantir a qualidade da pesquisa biomédica (BRUCE et al, 2016). No entanto, alguns resultados sugerem que a revisão editorial parece tornar os trabalhos mais legíveis e melhorar a qualidade geral dos relatórios (BRUCE et al, 2016). As intervenções avaliadas incluíam aspectos como o treinamento dos revisores, a adição de um estatístico no processo de revisão, o uso de listas de verificação, mudanças na modalidade da revisão por pares (identidade aberta ou não) e outras intervenções para aumentar a velocidade do processo de revisão. No entanto, como destacado em Grimaldo et al (2018), apesar de seu papel central na pesquisa, a revisão por pares tem sido examinada apenas por meio de projetos de pesquisa de pequena escala.

Quanto ao endosso de diretrizes de relato por parte das revistas, as evidências também são insuficientes para determinar o seu real impacto na completude dos relatórios de pesquisa (STEVENS, et al, 2014). No entanto, para algumas diretrizes, como o CONSORT e o TREND, o endosso por parte da revista parece estar associado a uma pequena melhoria na completude dos relatórios (TURNER et al, 2012; FULLER, et al, 2015). Em um estudo de escopo de revisões sistemáticas (n=50) sobre a adesão às diretrizes de relatórios (SAMAAN et al, 2013), os preditores da qualidade do relato de ensaios clínicos não randomizados incluíam o fator de impacto do periódico, a data de publicação (mais recente), a fonte de

financiamento (indústria), estudos de múltiplos centros, estudos com intervenções farmacológicas e o tamanho da amostra.

Outros aspectos da política editorial que podem estar relacionados à qualidade dos relatórios incluem as restrições de espaço e outras questões de ênfase estética (GINER-SOROLLA, 2012).

Bastuji-Garin et al (2013) realizaram um ensaio antes e depois em conjunto com análises de séries temporais com objetivo de avaliar o impacto da publicação do STROBE sobre a qualidade do relatório de estudos observacionais. O estudo incluiu 456 publicações de quatro periódicos dermatológicos de maior fator de impacto em 2010. Nas análises antes e depois, os autores observaram um aumento na completude dos relatórios (pontuação do STROBE) entre o período inicial pré-STROBE (2004–2005) e pós-STROBE (2008–2010), mas não entre o período imediato pré-STROBE (2006–2007) e 2008–2010. Por outro lado, nas análises de séries temporais interrompidas foi observado um aumento significativo na pontuação do STROBE ao longo do tempo. Os autores sugerem que as análises não controladas do tipo antes e depois, frequentemente utilizadas em pesquisas anteriores, podem superestimar os efeitos de intervenções como o uso de diretrizes de relato.

Podemos considerar as estruturas de incentivos mais amplas que operam na ciência como as características de um ecossistema (HIGGINSON & MUNAFÒ, 2016; SMALDINO & MCELREATH et al, 2016). Como sugerido em Smaldino & McElreath et al (2016), a seleção natural de certas práticas de pesquisa não requer, necessariamente, estratégias conscientes e trapaças por parte dos pesquisadores. Em vez disso, pode surgir da seleção positiva de métodos e hábitos que levam à publicação e ao maior sucesso em carreiras acadêmicas. Deste modo, práticas comuns em comunidades científicas podem mudar ao longo do tempo, não apenas, porque os pesquisadores estabelecidos estão estrategicamente mudando seus métodos e práticas, mas também porque certos pesquisadores são mais bem-sucedidos na transmissão de seus métodos para as gerações mais jovens (SMALDINO & MCELREATH et al, 2016). Neste contexto, os incentivos podem influenciar os padrões de inovação assim como a natureza da seleção.

3.2.2 Diferenças na atenuação da linguagem escrita

No início do estudo dos termos *hedges*, uma das primeiras hipóteses apresentadas pela linguista Robin Lakoff é que as mulheres seriam mais propensas do que os homens a utilizarem os recursos de atenuação da linguagem (LAKOFF, 1973). O estudo das diferenças de gênero na linguagem tem uma longa história, que abrange diferentes campos de pesquisa (gênero, linguística, psicologia). Em alguns contextos, a investigação de diferenças de gênero é considerada um fator importante, em outros contextos, pode ser considerado um tema controverso ou sem significado (KIMMEI, 2000)

No entanto, há poucas evidências disponíveis sobre o papel do gênero na expressão de incerteza no discurso escrito. Em uma meta-análise incluindo 29 (N=3.502) estudos sobre diferenças de gênero na linguagem tentativa falada, Leaper & Robnett (2011) observaram uma pequena diferença no uso de linguagem tentativa entre homens e mulheres. Foram considerados aspectos como a composição de gênero do grupo, a relação entre os parceiros de conversação, o status dos participantes, o tamanho do grupo, entre outros. Os autores sugerem que o uso de linguagem tentativa é altamente dependente do contexto (LEAPER & ROBNETT 2011).

A capacidade de expressar dúvidas e incertezas torna-se uma tarefa ainda mais complexa para escritores e aprendizes de segunda língua (HYLAND, 1996). Como destacado em Hinkel (1999), os escritores de segunda língua podem transferir os padrões discursivos típicos de sua língua e cultura de origem, podendo levar a falhas na aplicação de estratégias retóricas, dispositivos sintáticos e dos recursos de atenuação. Diferenças na atenuação da linguagem entre os escritores de primeira e segunda línguas foram observadas em estudos conduzidos em diferentes contextos (MIRZAPOUR & MAHAND, 2012; FARROKHI et al, 2008; KOURILOVA, 1995). Por exemplo, em um estudo de corpus (n=22800 palavras) de artigos científicos de nativos de língua inglesa, a frequência média observada de termos *hedges* (por 100 palavras) foi de 6,5 nas seções de Introdução e Discussão, enquanto, entre os artigos dos não nativos de língua inglesa a frequência média de *hedges* foi de apenas 2,1 (KOURILOVA, 1995). Além disso, foi observado que aproximadamente

80% dos termos *hedges* utilizados entre os falantes não nativos referiam-se apenas ao verbo modal *may* (KOURILOVA, 1995).

4 APLICAÇÕES E CARACTERÍSTICAS DOS DIAGRAMAS CAUSAIS

4.1 Introdução

Nas últimas décadas, os diagramas causais têm sido apontados como uma das principais ferramentas que podem contribuir para a qualidade metodológica e do relato de estudos observacionais (PETERSEN & VAN DER LANN, 2014; PORTA & BOLÚMAR, 2015; GLYMOUR, 2006). Essas ferramentas têm sido utilizadas para demonstrar em quais situações o uso de critérios e técnicas tradicionais para tratar confundimento, viés de seleção e informação pode levar a conclusões inadequadas (HERNÁN et al., 2002; 2004; ROBINS, 2001). Além disso, permitiram compreender algumas controvérsias presentes na literatura epidemiológica, como a natureza do viés de Berkson e do Paradoxo de Simpson (PEARL, 2013). No entanto, apesar do crescente reconhecimento do seu potencial, ainda não há uma descrição sobre como os diagramas causais têm sido utilizados nas investigações empíricas.

Neste capítulo foi realizada uma revisão da literatura com o objetivo de descrever o quanto, como e onde os diagramas causais têm sido utilizados em estudos observacionais analíticos nos últimos 18 anos. Foram analisados os itens relativos ao crescimento anual das publicações, os principais periódicos, padrões de autoria e coocorrência de palavras chaves. Além disso, foram avaliadas as informações da conectividade dos DAG, a representação de aspectos relacionados às limitações do estudo e o relato de itens importantes na comunicação de confundimento.

4.2 Métodos

4.2.1 Seleção dos estudos

A estratégia de busca visou a identificação de publicações de estudos observacionais analíticos que utilizaram as ferramentas dos diagramas causais. A “utilização” de diagramas causais foi definida como a apresentação no artigo (ou no material suplementar) de um diagrama causal na forma de um gráfico acíclico direcionado (DAG) ou o relato da utilização dessas ferramentas em qualquer fase do estudo. Foram incluídas as publicações de estudos observacionais cujo objetivo foi investigar a relação entre uma (ou mais) variáveis de exposição e de desfecho, utilizando um grupo de comparação.

Foram excluídas as publicações de (i) estudos metodológicos com ou sem aplicação empírica: em geral, estudos envolvendo demonstrações de uma ferramenta ou método; (ii) estudos de simulação sem aplicação empírica; (iii) estudos não observacionais (iv) cartas, resumos de congressos, artigos de revisão e editoriais; (v) publicações com versões paramétricas de gráficos acíclicos direcionados, incluindo as aplicações em análise fatorial e análise de caminhos; (vi) publicações que utilizaram outros tipos de modelos gráficos (*Partial-DAG, bayesian networks, influence diagrams, causal loop diagrams, conceptual/map causal networks*); (vii) publicações referentes à descoberta de estruturas causais ou sobre temas não relacionados à inferência causal; e (viii) artigos publicados em outras línguas além do inglês.

A busca eletrônica foi conduzida em duas bases de dados: Medline (via Pubmed) e Web of Science (WOS – Thomson Reuters). O período de publicação foi limitado a 18 anos, 2000-2017. As chaves de buscas foram construídas a partir de variações dos seguintes descritores: *causal DAG; directed acyclic graph; causal diagram; causal graph; structural causal model; causal graphical model; causal network*.

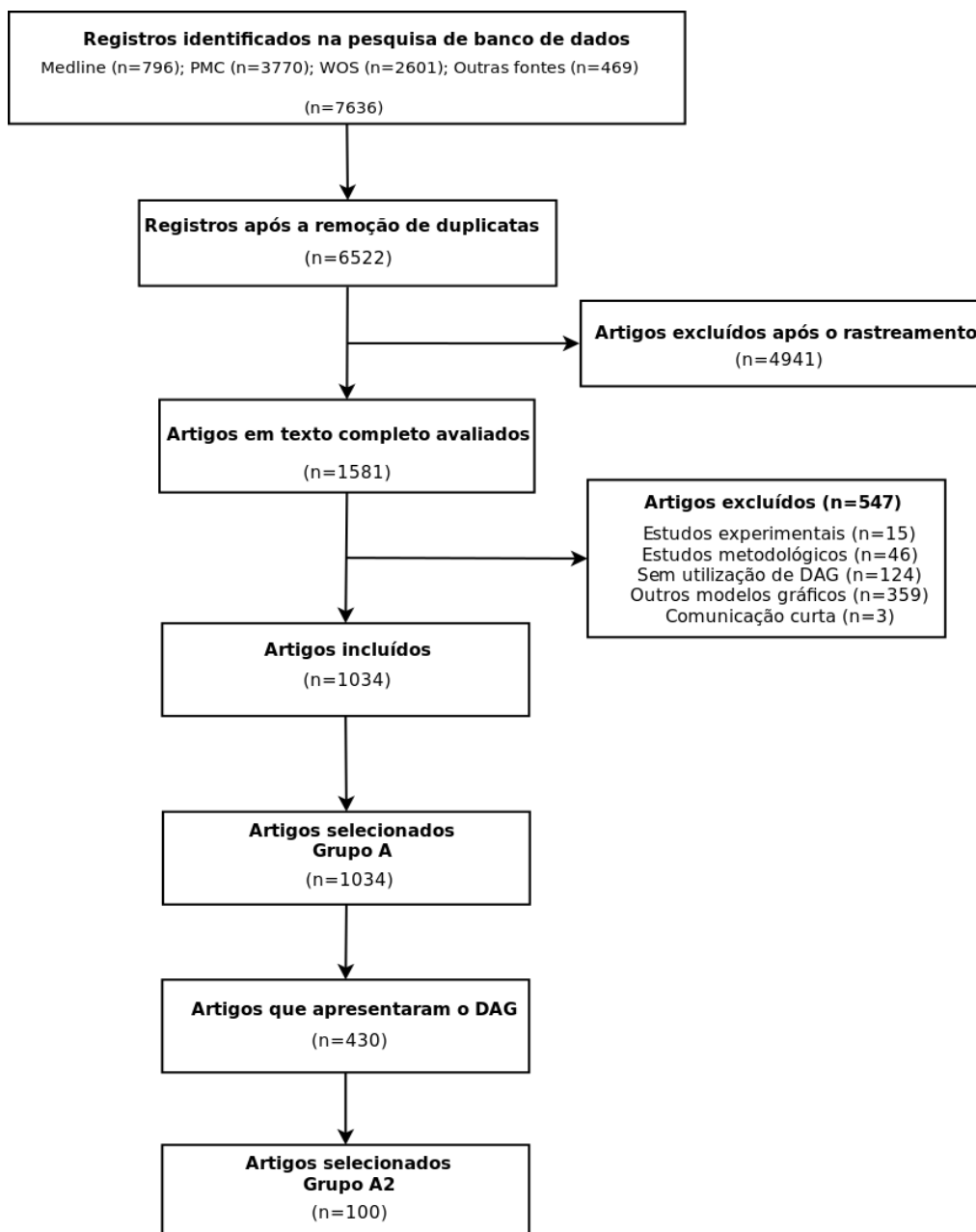
Dado que os termos de interesse podem estar presentes apenas no texto completo e nos títulos das figuras, foi realizada uma busca no repositório digital PubMed Central (PMC). As estratégias de busca completas estão descritas no APÊNDICE A. Em uma segunda etapa, foi realizada uma busca por citações por meio da base de dados da Web of Science (Thomson Reuters). Na pesquisa por citações foram consideradas 2 publicações metodológicas altamente citadas entre

os artigos selecionados na primeira etapa (TEXTOR et al, 2011; SHRIER & PLATT, 2008)

Após a execução das estratégias de busca, os registros identificados (títulos e/ou resumos disponíveis) foram reunidos em um banco de dados de um gerenciador de referência (EndNote). Após a exclusão de registros duplicados, os títulos e resumos dos demais artigos foram avaliados pelo revisor principal (TC) considerando os critérios elegibilidade. Nesta fase, os registros foram classificados em três categorias: “incluído”, “excluído” ou “incerto”. Os registros considerados “incluídos” ou “incertos” passaram automaticamente para o próximo nível de rastreio (nível 2). Posteriormente, os textos completos de todos os artigos potencialmente elegíveis foram reavaliados pelo revisor principal quanto a elegibilidade. Todos os registros considerados “excluídos” foram reavaliados por meio de busca automática de termos.

A pesquisa nos bancos de dados produziu 7636 registros potencialmente relevantes (3770 PMC; 2601 WOS; 796 Medline; 469 fontes adicionais). Após as exclusões de duplicatas e rastreamento por meio do título e resumo, 1581 publicações permaneceram elegíveis para revisão de texto completo. Foram excluídas 531 publicações: 46 estudos com objetivos metodológicos, 124 publicações sem a utilização de diagramas causais, 359 artigos que utilizaram outros tipos de modelos gráficos, 15 publicações de estudos experimentais com ou sem randomização e 3 publicações no formato de comunicação curta (Figura 9). Para as análises bibliométricas foram considerados todos os 1034 artigos identificados nas buscas nas bases de dados e selecionados de acordo com os critérios de elegibilidade (Grupo A). Para avaliação das características dos diagramas causais e do relato de confundimento, foi selecionada uma amostra aleatória de N=100 (Grupo A2), obtida a partir do total de artigos que apresentaram a estrutura gráfica no artigo ou no material suplementar (N=430).

Figura 9 Fluxograma da seleção dos estudos.



Fonte: CORTES, 2018.

4.2.2 Coleta e análise de dados

4.2.2.1 Análises bibliométricas

Todos os artigos selecionados (Grupo A) foram analisados quanto ao ano de publicação, afiliação dos autores, principais periódicos, padrões de citações e altimetria e coocorrência de palavras chaves nos títulos e resumos.

Os dados foram apresentados como o número de registros recuperados em cada ano em todas as bases de dados utilizadas. Para a normalização das contagens anuais foram considerados apenas os dados do Medline. Foram realizados dois tipos de comparações: o número absoluto de publicações com a utilização de diagramas causais foi dividido (i) pelo total de artigos indexados por ano no Medline e (ii) pelo total de artigos identificados por meio de termos de busca para estudos etiológicos (APÊNDICE A). Apesar dos registros recuperados corresponderem às diferentes bases de dados, 99% dos artigos selecionados estão indexados no Medline, 97% na WOS e 85% no PMC.

Para determinar o país de afiliação do primeiro autor e todos os coautores, os endereços institucionais completos de cada coautor foram processados no ambiente R. Todos os artigos que incluíram autores com endereços de diferentes países foram codificados como colaborações internacionais. Para facilitar o processamento dos dados e a verificação de inconsistências, as informações de autoria e de fontes de publicação foram padronizadas a partir dos dados da Web of Science. Para 34 publicações, que não estão indexadas na WOS, foram utilizados os dados do Medline.

Na avaliação do nível de “atenção acadêmica” do conjunto de publicações, foram consideradas duas métricas tradicionais: o número de citações e o índice H (BARRETO et al, 2013). Os dados sobre o índice H e a frequência das citações foram obtidos por meio da ISI Web of Science. Para cada artigo da amostra, foi coletado o número de citações recebidas até 31 de dezembro de 2017. Isso implica

que há diferentes janelas de citações, dependendo do ano de publicação do artigo, e o “atraso de citação” para os artigos mais recentes torna a análise de citações menos informativa.

Para complementar as métricas de citações, foi selecionada uma métrica alternativa (altimétrica) de atenção online. As informações sobre o *Altmetric Attention Score* foram obtidas do banco de dados da Altmetric LLP por meio da biblioteca *rAltmetric* do R. Essa pontuação pode ser interpretada como um indicador da quantidade de repercussão online de cada publicação. A pontuação altimétrica inclui uma medida ponderada das menções do artigo em mídias sociais, serviços de bookmarking acadêmico, notícias científicas, fóruns pós-publicação e outras fontes on-line (Wikipedia, Youtube, Reddit/Pinterest, LinkedIn, etc) (TRUEGER et al, 2015). Adicionalmente, foram coletadas as informações sobre o contexto do escore em termos de percentis. Isto é, o percentil de cada artigo considerando as estatísticas dos artigos em todo o banco de dados (Contexto Geral), dos artigos publicados na mesma revista (Contexto Periódico), dos artigos publicados no mesmo período – 6 semanas para mais e para menos (Contexto Idade da publicação) e dos artigos da mesma revista e idade (Contexto Periódico e Idade da publicação).

A análise de coocorrência de palavras chaves foi realizada no programa de livre acesso VOSviewer (www.vosviewer.com). Detalhes sobre a técnica VOS utilizada na normalização, e o mapeamento e definição de clusters são descritos em Van Eck et al (2010). De forma geral, cada termo (vértice) é atribuído a grupo (cluster) e a proximidade entre os vértices indica a frequência de coocorrência entre os termos. A força de associação (tamanho das setas) entre os vértices é ponderada pela frequência do termo. Para a construção da rede, foi utilizado como critério a ocorrência mínima do termo em 10 artigos. Além disso, foi criado um arquivo tesouro para agregar os termos com significados semelhantes, incluindo sinônimos, formas plurais e singulares e abreviaturas. Os detalhes da padronização e exclusão de palavras são apresentados no APÊNDICE B. Para a criação do tesouro foi utilizada a interface do wordnet (<https://wordnet.princeton.edu/>) no ambiente R.

4.2.2.2 Características dos diagramas causais

Para a amostra de artigos que apresentaram a estrutura gráfica (Grupo A2, n=100), foram extraídos as informações sobre o tipo de aplicação e o número de modelos gráficos apresentados. Foi avaliado se os diferentes DAG se referem à mesma pergunta de pesquisa ou se um único DAG foi utilizado para guiar as análises de diferentes exposições e/ou desfechos.

Diagramas causais com desenho de estudo (DAG): foi investigado se alguns dos modelos gráficos incluíam aspectos do desenho de estudo, como a presença de vértices (explicitamente) indexados no tempo ou vértices para indicar aspectos de mensuração ou seleção. Destaca-se que o termo “DAG com desenho de estudo”, sugerido em Karvanen & Finland (2014), foi utilizado apenas para a simplificação da descrição dos modelos gráficos com amostragem explícita ou mecanismo de seleção, variáveis não observadas, etc.

Relato de incerteza sobre a estrutura causal: foi avaliado se os autores apresentaram algum referencial teórico para o modelo causal e o relato de fontes de incerteza sobre a estrutura gráfica. Nesses itens, foram considerados apenas os comentários direcionados à estrutura gráfica.

Características relacionadas à conectividade do gráfico: foram extraídos as informações sobre o número de vértices e o número de arestas. Também foi verificada a presença, o número e o tipo de variáveis agrupadas em um único vértice.

A densidade do DAG foi calculada como o número de arestas do gráfico dividido pelo número máximo de arestas (MA) de um DAG: $MA = (n-1)(n)/2$, onde n é o número de arestas do gráfico. Um gráfico com apenas algumas arestas, é chamado de gráfico esparsos. Um gráfico denso é um gráfico em que o número de arestas está próximo do número máximo de arestas. A definição para esses dois termos não é muito precisa, ou seja, não há um limite superior para uma densidade esparsa e nenhum limite inferior (mínimo) para definir um gráfico denso. A

densidade é 0.5 para um gráfico com metade do número máximo de setas e 1 para um gráfico completo.

4.2.2.3 Relato de confundimento

Para os artigos da amostra (Grupo A2, n=100) foram extraídas as informações relativas à concepção e análise de confundimento. Os itens sobre o relato de confundimento foram selecionados tendo como base a declaração de diretrizes para o relato de estudos observacionais (STROBE – VON ELM et al, 2007) e outras publicações sobre o tema (GROENWOLD et al, 2008; POUWELS et al, 2016). A extração de dados foi realizada por apenas um revisor (TC). A versão final do questionário é apresentada no APÊNDICE B. Foram extraídas as informações sobre os seguintes itens:

(1) Se os autores relataram as razões pelas quais os potenciais fatores de confusão foram selecionados. Este item não faz parte das recomendações do STROBE, mas é um aspecto importante na avaliação de confundimento e permitirá a identificação dos estudos que relataram apenas a utilização de critérios teóricos ou que utilizaram critérios de seleção adicionais.

2) Se os autores relataram as razões para a inclusão dos fatores de confusão no modelo final. O item 16 do STROBE recomenda que os autores descrevam quais foram os confundidores utilizados no ajuste e porque foram incluídos no modelo. Com base nos critérios relatados nos itens 1 e 2, as estratégias de seleção para o modelo final foram classificadas como: a) apenas critérios teóricos (DAG – modelo total); b) Critérios teóricos e critérios estatísticos; c) apenas critérios estatísticos.

(3) Se os autores relataram as características dos principais fatores de confusão. O item 7 do STROBE recomenda que os autores apresentem as definições das variáveis, incluindo os confundidores em potencial. O item 14^a (STROBE) recomenda a apresentação das características dos participantes, incluindo as informações dos confundidores.

(4) Se os autores relataram o método estatístico utilizado para o ajuste de confundimento (STROBE-12a). Além disso, foram extraídas as informações sobre os tipos de métodos utilizados.

(5) A apresentação das estimativas de efeito bruto e também das estimativas de efeito ajustado (STROBE-16a). Quando apenas as estimativas ajustadas foram relatadas, o item foi classificado como “sim”, se estão disponíveis no artigo ou no material suplementar os dados para o cálculo das estimativas brutas.

(6) Se os autores reconheceram a possibilidade de confundimento não observado/residual e se discutiram qualitativamente a possível direção do viés de confundimento.

(7) O item 12e do STROBE recomenda a descrição de qualquer análise de sensibilidade. Neste item foi considerado qualquer tipo de análise descrita pelos autores com o objetivo de avaliar o impacto do confundimento. As análises foram classificadas como **a) técnicas de modelagem de viés**: foram incluídas nessa categoria as técnicas que utilizam uma distribuição dos parâmetros do viés para reajustar as medidas de efeito e precisão (*Probabilistic bias analysis, Bayesian sensitivity analysis, Monte Carlo sensitivity analysis*) e **b) técnicas não probabilísticas (fórmulas de viés)**: que permitem calcular algebricamente a magnitude do viés necessária para explicar a associação observada e **c) outros tipos de análise de sensibilidade**: em geral, estratégias envolvendo a inclusão ou exclusão de variáveis do modelo, restrição da amostra e a redefinição de variáveis.

Além dos itens derivados do STROBE, foram extraídas as informações sobre o número de variáveis de confusão incluídas no modelo final. Para os artigos que relataram apenas estratégias teóricas de seleção dos confundidores (isto é, DAG – modelo total), foi avaliada a ocorrência de modificações não relatadas no conjunto suficiente de ajuste. Ou seja, o conjunto de confundidores incluídos no modelo final foi comparado com o conjunto suficiente derivado do DAG apresentado no artigo.

Outro aspecto investigado foi a possibilidade da “falácia da tabela 2” (*table 2 fallacy*, WESTREICH & GREENLAND, 2013). A possibilidade da “falácia da tabela 2” foi definida como a apresentação de todos os coeficientes estimados a partir de um único modelo de regressão. Além disso, foi avaliado se a) os autores interpretaram essas estimativas e b) se fizeram alguma distinção nas interpretações.

4.2.2.4 Análises de dados

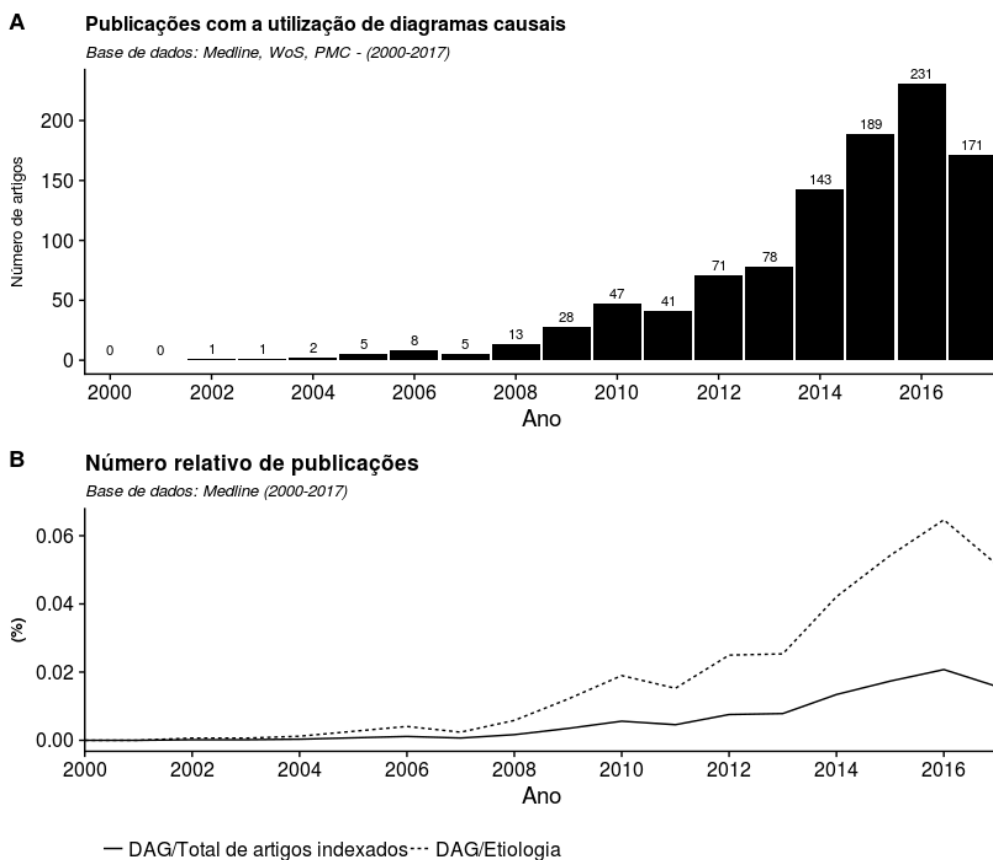
As variáveis foram descritas com frequências e porcentagens (%) ou medianas e intervalos interquartis (IQR). Os resultados sobre do relato de confundimento foram apresentados como porcentagem nos itens individuais. As análises estatísticas foram conduzidas no ambiente R (versão 3.0.2). Todos os pacotes de linguagem R utilizados nas análises são descritos nos scripts apresentados no APÊNDICE 3. O arquivo de fonte de dados bruto e os scripts R serão armazenados em um repositório de hospedagem *Open Science framework* (<https://osf.io/>).

4.3 Resultados

4.3.1 Análises bibliométricas das publicações com diagramas causais

Para o período de 2000 a 2017, foram identificadas 1034 publicações de estudos observacionais analíticos que utilizaram as ferramentas dos diagramas causais. A figura 1A apresenta o número de artigos recuperados por ano em todas as bases de dados avaliadas. O primeiro artigo identificado foi publicado em 2002. Em 2016, foi registrado o maior número de publicações (n=222 artigos, 21,5%). Os artigos indexados cresceram 41% ao ano durante o período de estudo. Em comparação com os artigos indexados com descritores para estudos etiológicos, a contribuição relativa dos artigos com a utilização de diagramas causais aumentou moderadamente, atingindo 0,06% em 2016 (figura 1B).

Gráfico 1. Número de publicações por ano de estudos observacionais analíticos com o relato da utilização de diagramas causais.



Legenda: (a) número de publicações por ano de estudos observacionais analíticos com o relato da utilização de diagramas causais. (b) número de publicações por ano normalizado pelo total de artigos indexados no medline e pelo total de artigos com termos relacionados à inferência causal.

Fonte: CORTES, 2018.

As publicações identificadas totalizaram 5021 autores e 85 países de afiliação. O índice de colaboração internacional foi de 41%. A principal origem de afiliação dos primeiros autores foram os EUA ($n = 553$, 53%), seguido da Inglaterra ($n=52$, 5%), Noruega ($n=49$, 4,7%), Alemanha ($n=46$, 4,4%), Suécia ($n=41$, 3,9%), Dinamarca ($n=40$, 3,8%), Austrália ($n=36$, 3,5%), Canadá ($n=25$, 2,4%), Holanda ($n=20$, 1,9%) e Espanha ($n=16$, 1,5%) (Tabela 3).

Tabela 3. Principais países de afiliação dos coautores de publicações com o relato da utilização de diagramas causais.

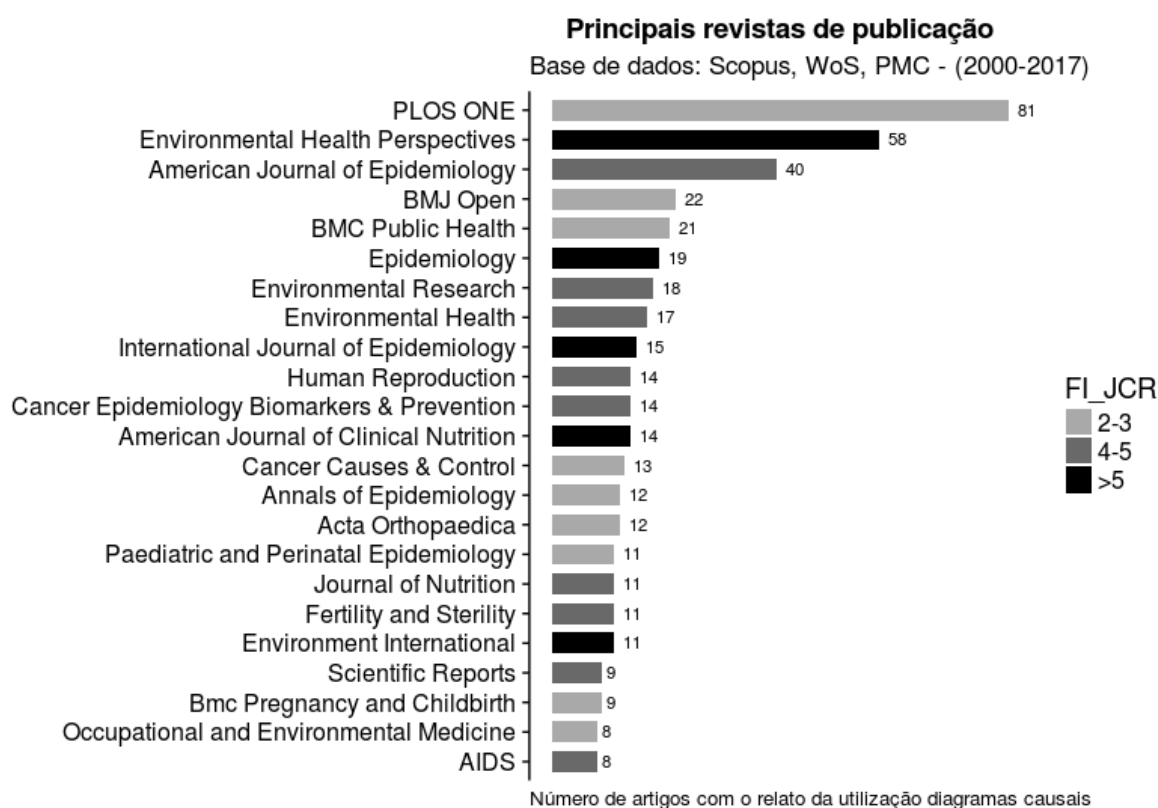
País	Primeiro autor N = 1034		Todos os coautores N = 5021	
	N	(%)	N	(%)
Estados Unidos	553	53	683	13
Inglaterra	52	5,0	120	2,9
Noruega	49	4,7	75	1,5
Alemanha	46	4,4	71	1,4
Suécia	41	3,9	79	1,6
Dinamarca	40	3,8	85	1,7
Austrália	36	3,5	49	1,0
Canadá	25	2,4	76	1,5
Holanda	20	1,9	50	1,0
Espanha	16	1,5	41	0,8
França	9	0,9	37	0,7
Brasil	9	0,9	23	0,5

Nota: Base de dados Web of Science, Medline, PMC; Período 2000-2017
Fonte: CORTES, 2018.

As três principais fontes de publicação foram as revistas *PLOS ONE* (7,8%, n=81), *Environmental health perspectives* (5,6%, n=58) e *American Journal of Epidemiology* (3,9%, n=40) (Gráfico 2). A mediana do fator de impacto (JCR – 5 anos) das 23 principais revistas foi de 12 (6,5-17,5).

Os artigos recuperados receberam um total de 10.382 citações, 692.13 citações por ano e com uma mediana de 9 citações por artigos (4-21). O h-index foi de 47. Mais da metade dos artigos (n=441, 74%) recebeu pelo menos uma menção em notícias on-line, blogs científicos ou mídias sociais. A mediana para o escore de atenção online foi de 1,85 (intervalo, 0,25-9,10). 37% (n=379) das publicações estão no percentil 80 no contexto geral, e 32% (n=327) estão no percentil 80 no contexto da mesma idade e revista. A maior pontuação *Altmetric* foi de 2.539 (<https://www.altmetric.com/details/2819326/news>).

Gráfico 2. Principais fontes de publicação dos artigos com diagramas causais.



Fonte: CORTES, 2018.

O mapeamento de palavras-chave utilizadas nos títulos e resumos dos artigos com a utilização de diagramas causais é apresentado na figura 10. As redes de coocorrência de palavras chaves podem auxiliar na identificação dos principais tópicos de pesquisa de um grupo de publicações. O mapeamento pressupõe que, quanto mais frequente a coocorrência de um par de palavras, mais semelhantes são os temas que eles indicam. Deste modo, os clusters representados por diferentes cores podem ser interpretados como temas de pesquisa. Considerando como critério a ocorrência mínima do termo em 10 artigos e após a padronização e exclusão de palavras, 190 termos foram selecionados e classificados em quatro clusters. No cluster relacionado à saúde materno-infantil (em roxo), variações do termo *child* ocorreram 271 vezes no título ou no resumo dos artigos, representando 26% das publicações. Outros termos relacionados foram *pregancy* (226 vezes), *birth* (150) e *mother* (114). No cluster em azul, os principais termos foram *obesity* (205), *race*

4.3.2 Características dos diagramas causais

Na tabela 4 são apresentadas as principais características dos diagramas causais incluídos nos artigos da amostra (N=100). Os DAG foram utilizados principalmente na seleção de covariáveis para as análises de efeito total (n=71). Outras aplicações envolveram análises de mediação (n=17) e análises por variáveis instrumentais (n=3). Nove artigos utilizaram os modelos causais gráficos apenas na discussão das hipóteses ou resultados.

Apenas 23 publicações relataram algum referencial teórico para o modelo gráfico. Dentre essas, apenas 3 artigos apresentaram algum referencial teórico para a exclusão de setas. O relato de fontes de incerteza sobre a estrutura gráfica também foi pouco frequente (n=24). A maioria dos artigos que mencionou incertezas sobre a presença de variáveis ou setas (n=21), sobre a direção das relações (n=11) ou sobre o papel de uma variável (n=6), discutiram como esses aspectos poderiam afetar as conclusões derivadas do modelo causal.

Mais da metade dos diagramas causais contém algum aspecto do desenho de estudo (n=51). A presença de vértices não observados foi o aspecto mais prevalente (n=38). Dentre esses, 23 modelos gráficos apresentam ancestrais não observados da exposição e do desfecho e 4 apresentam causas comuns não observadas do mediador e do desfecho.

A maioria das publicações contém apenas um DAG causal (n=87). Apenas 3 publicações utilizaram mais de uma estrutura gráfica para guiar diferentes análises para a mesma pergunta de pesquisa. Quatro artigos utilizaram diferentes estruturas para discutir hipóteses alternativas para a interpretação dos resultados.

A mediana de vértices e arestas foi 9 (intervalo, 6-12) e 18 (intervalo 10-29), respectivamente. A mediana da densidade foi 0,56 (intervalo, 0,37-0,75) e 8 DAG são gráficos completos.

Tabela 4. Aplicações e características dos diagramas causais (N=100).

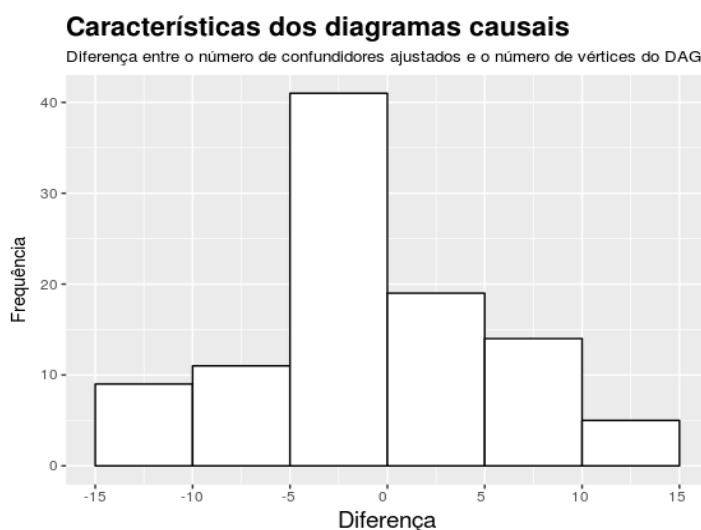
Item	n
Apresentação do DAG	
No artigo	66
No material suplementar	33
Aplicação do DAG	
Análises dos dados	91
Apenas na discussão das hipóteses ou resultados	9
Tipo de análise	
Efeito total	70
Análise de mediação	16
Análise por variáveis instrumentais	3
Único DAG para múltiplas perguntas de pesquisa	
Sim	33
Não	67
Múltiplos DAG para mesma pergunta de pesquisa	
Sim	7
Não	93
Apresentação do referencial do DAG	
Comentários com referências	23
Comentários gerais sobre o DAG	25
Nenhum	52
DAG com desenho de estudo	
Sim	51
Não	49
Aspecto do desenho de estudo*	
Vértices não observados	38
Vértices indexados no tempo	16
Outros aspectos de mensuração	6
Aspectos de seleção	3
Relato de incerteza sobre a estrutura gráfica	
Sim	24
Não	76
Tipo de incerteza*	
Presença de variáveis ou setas	21
Direção das relações	11
Papel de uma variável	6
Presença de variáveis agrupadas	
Sim	59
Não	41
Tipos de vértices agrupados*	
Exposições ou desfechos	14
Mediadores	13
Ancestrais da exposição/desfecho	42

Nota: *Itens com mais de uma opção de resposta.

Fonte: CORTES, 2018.

No entanto, mais da metade dos modelos gráficos apresenta variáveis agrupadas em um único vértice (n=59). A mediana de vértices agrupados foi 2 (IQR, 1-4). O principal tipo de vértice com variáveis agrupadas corresponde aos ancestrais da exposição e do desfecho (n=42). A figura 11 apresenta a diferença entre o número de variáveis de confusão incluídas no modelo e o número de vértices do DAG (excluindo a exposição e o desfecho). Na ausência de variáveis agrupadas, espera-se que o número de confundidores ajustados seja menor que o número de vértices do gráfico (dado a exclusão de E e D).

Figura 11. Diferença entre o número de variáveis de confusão incluídas no modelo final e o número de vértices do DAG.



Legenda: Valores positivos indicam o número de variáveis agrupadas.

Fonte: a CORTES, 2018.

4.3.3 Relato de confundimento

Na Tabela 5 são apresentadas as frequências do relato de itens importantes na comunicação de confundimento nos artigos da amostra (n=100). Apenas 7

publicações não mencionaram nenhum critério de seleção dos potenciais fatores de confusão e três publicações não realizaram ajuste de confundimento, utilizando o DAG apenas na seleção dos potenciais mediadores.

Na seleção dos confundidores para o modelo final, a maioria dos artigos (n=72) mencionou apenas a utilização do DAG (DAG – modelo total). 19 publicações relataram modificações no conjunto de ajuste, incluindo critérios adicionais de significância estatística (n=6), critérios de mudanças nas estimativas (n=10) e critérios de precisão (n=3). Apenas três publicações realizaram modificações não relatadas no conjunto suficiente derivado do DAG apresentado no artigo. Além disso, poucas publicações (n=3) mencionaram alguma informação sobre o número de conjuntos suficientes.

Vinte publicações foram classificadas como possível ocorrência da falácia da tabela 5. Apenas quatro artigos mencionaram possíveis diferenças na interpretação das estimativas.

Outros itens que foram mais frequentemente relatados foram os métodos utilizados para o ajuste de confundimento (n=98), as características dos principais fatores de confusão (n=96) e apresentação de ambas estimativas ajustadas e não ajustadas (n=87). Mais da metade das publicações reconheceram a possibilidade de confundimento não observado (n=60). Entretanto, os relatórios incluíram menos frequentemente as declarações qualitativas sobre a possível direção do viés (n=19) e o uso de análises de sensibilidade para a avaliação do potencial impacto do confundimento não observado (n=31).

Tabela 5. Frequência de itens importantes na comunicação de confundimento em uma amostra de artigos com diagramas causais (N=100).

Itens	n
Relato das razões para a seleção dos potenciais fatores de confusão	90
Sem ajuste de confundimento	3
Relato das razões para a inclusão dos fatores de confusão no modelo final*	93
DAG – Modelo total	72
Outros critérios teóricos	3
Critérios de mudanças na estimativa	10

Critérios de significância estatística	6
Critérios de precisão	3
Modificações não relatadas no conjunto de ajuste	3
Relato do método utilizado para o controle de confundimento*	98
Regressão de riscos proporcionais de Cox	12
GLM	56
GEE	7
GLMM	5
Modelo multinível	5
Modelo marginal estrutural	6
Escore de propensão	1
Variável instrumental	3
Outros	10
Relato das características dos principais fatores de confusão*	96
Relato das estimativas de efeitos não ajustadas e ajustadas*	87
Possibilidade de falácia da tabela 2	20
Sem interpretação das estimativas	7
Mesma interpretação	9
Relataram possíveis diferenças	4
Reconheceram a possibilidade de confundimento não observado*	60
Declaração qualitativa sobre a possível direção do viés de confusão*	19
Análises de sensibilidade para a avaliação de confundimento não observado*	31
Modelagem de viés	2
Fórmula de viés	3
Outros tipos de análise de sensibilidade	26

Nota: * Itens presentes no STROBE.

Fonte: CORTES, 2018.

4.4 Discussão

Neste trabalho foi realizada uma revisão da literatura em conjunto com ferramentas bibliométricas com o objetivo de descrever o quanto, como e onde os diagramas causais têm sido utilizados nos estudos observacionais analíticos nos últimos 18 anos.

Os resultados mostram uma tendência de crescimento no número de aplicações dessas ferramentas. Foi observado um crescimento mais acentuado após o ano de 2010, possivelmente devido à maior disponibilidade de ferramentas

para a construção e análise de diagramas causais. Nesse período, foram lançadas a ferramenta de web, *DAGitty* (TEXTOR, 2011), o programa em DOS, *DAG program* (KNÜPPEL, 2010), e um conjunto de funções em R posteriormente incorporadas ao pacote *dagR* (BREITLING, 2010). Outras formas de incentivo à aplicação da teoria dos diagramas causais, incluem as diversas publicações metodológicas sobre o tema, tutoriais publicados em outras línguas além do inglês (Japonês: SUZUKI et al, 2009; Alemão: SCHIPF et al, 2011; Francês: DUMAS et al, 2014; Espanhol: SÁNCHEZ-NIUBÒ, et al, 2016; Português: CORTES et al, 2016; Chinês: ZHENG et al, 2017) e prêmios de incentivo ao ensino como o *Causality in Statistics Education award* (<http://www.amstat.org>).

Considerando as escolhas e variações nas bases de dados e questões como a localização das palavras chaves nos artigos, os registros identificados podem representar apenas uma parcela das publicações com a utilização de diagramas causais. Além disso, esta revisão não incluiu artigos publicados em outras línguas além do inglês e outros tipos de publicações como teses de mestrado e doutorado. Os resultados sobre fontes de publicação também são influenciados por alguns desses aspectos, especialmente os resultados das buscas em texto completo (PMC), devido às variações no tipo e no nível de participação das revistas no repositório digital PMC.

Além disso, os padrões de publicação foram utilizados para avaliar o nível de disseminação dos diagramas causais em estudos observacionais. No entanto, o impacto de uma teoria inclui outros aspectos não abrangidos na contagem de publicações. Como destacado em Porta (2000), o surgimento de novos conceitos, a resolução de paradoxos (PEARL, 2013) e a expansão dos métodos de identificação (PEARL, 2010; PEARL & BAREINBOIM, 2014) são alguns dos indicadores da possível “revolução” ou renovação metodológica impulsionada pela teoria dos diagramas causais.

Nas próximas seções são discutidos os resultados sobre as características dos DAG e os aspectos do relato de confundimento. Além disso, são apresentadas algumas recomendações gerais para a construção, análise e apresentação de DAG causais em estudos observacionais.

4.4.1 Construção e apresentação do DAG causal

Neste trabalho, foram avaliadas as informações relacionadas à conectividade dos DAG, a representação de aspectos do desenho de estudo e a comunicação explícita de fontes de incertezas sobre a estrutura causal. Em conjunto, essas informações podem indicar a “força” das hipóteses causais codificadas nos DAG e o nível de incerteza dos autores sobre os modelos apresentados.

Metade dos diagramas causais avaliados incluem algum aspecto do desenho de estudo (51%). No entanto, poucos modelos gráficos contêm ancestrais não observados da exposição e do desfecho (confundidores) (n=23), ou a representação de erros de mensuração (n=6) e mecanismos de seleção (n=3). Além disso, o relato de incertezas sobre a estrutura causal também foi pouco frequente (n=24).

Uma possível interpretação é que a utilização de DAG causais ou apresentação do modelo gráfico no artigo seja condicionada ao nível de incerteza dos autores sobre o tema de pesquisa e sobre a qualidade metodológica do estudo. No entanto, a metodologia de um estudo nem sempre corresponde aos fenômenos complexos que dão origem aos dados.

Alguns autores já reforçaram a importância da construção de modelos causais “realistas” que representam simultaneamente as fontes comuns de viés em estudos epidemiológicos (GREENLAND, 2010; HERNÁN & ROBINS, 2016). Detalhes sobre a representação de erros de mensuração são apresentados em Hernán & Cole (2009). A representação explícita do mecanismo de seleção e estratégias de identificação são apresentados em Didelez et al (2010), Bareinboim e Pearl (2012), Karvanen (2016).

Em alguns casos a necessidade de relaxamento das hipóteses causais implicará em não identificação do efeito causal. Neste contexto, diferentes estratégias podem ser adotadas para lidar com tensão entre hipóteses mais “realistas” e a identificação do parâmetro causal. Primeiro, como destacado em Petersen & Van der Laan (2014), é importante diferenciar rigorosamente as hipóteses que representam o conhecimento real sobre os processos que geraram os dados (*knowledge-based assumptions*), das demais hipóteses que, se fossem

verdadeiras, resultariam em identificação do parâmetro causal (*convenience-based assumptions*).

Uma possível estratégia consiste na especificação e apresentação de múltiplos modelos gráficos. Esses modelos (ou um subconjunto) podem ser utilizados para guiar diferentes análises e apresentados com as estimativas de efeito correspondentes (EVANS et al, 2012; LASH et al, 2014) ou podem ser utilizados para se obter uma estimativa média ao longo de vários modelos candidatos (HAMRA et al, 2015). Mesmo que análises adicionais não possam ser conduzidas (p. ex, por meio análises de sensibilidade), os modelos candidatos podem ser utilizados na discussão das hipóteses alternativas para a interpretação dos resultados. Essas estratégias podem contribuir para uma comunicação mais completa das fontes de incerteza e para o direcionamento de pesquisas futuras. Além disso, podem minimizar os vieses decorrentes da apresentação seletiva dos modelos (p. ex. *wish bias* ou *white hat bias*) (COPE, 2010; HAMRA et al, 2015).

Considerando a mediana de vértices e arestas dos DAG da amostra, são apresentados, na Figura 5, quatro DAG com diferentes estruturas de adjacência e de colisão (imoralidades). Nota-se que as diferentes hipóteses causais representadas nos DAG A e B implicam as mesmas conclusões sobre o conjunto de ajuste para o efeito total de E em D (S: {C1, C4, C5, C6}). No entanto, todos os 4 DAG implicam em diferentes relações de independência condicional e, portanto, podem ser distinguidos a partir dos dados de estudos observacionais (GREENLAND & PEARL, 2010). Em outros casos, diferentes estruturas causais podem implicar o mesmo conjunto de restrições sobre os dados observados. Ou seja, podem ser estatisticamente equivalentes.

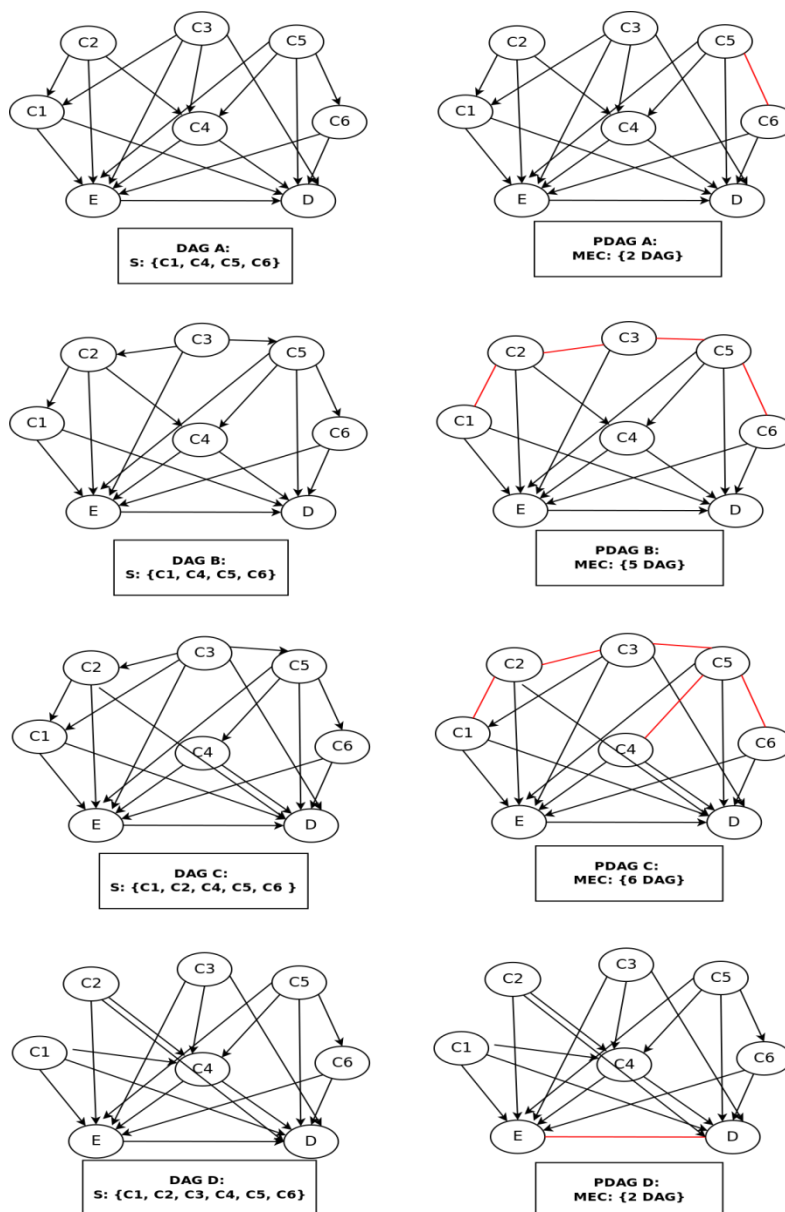
Na figura 13, a classe de modelos (markov) equivalentes de cada DAG é representada por um gráfico parcialmente direcionado (*complete partial directed acyclic graph* – cPDAG ou *essencial graph*) (RADHAKRISHNAN et al, 2017). Em um cPDAG, as arestas direcionadas têm a mesma orientação em todos DAG da classe e as arestas não direcionadas (em vermelho) podem ser orientadas de qualquer forma, desde que não crie um ciclo ou uma nova estrutura de colisão. O tamanho de uma classe de equivalência de markov (MEC) é o número de DAG pertencentes a classe. Neste exemplo, considerando os mesmos vértices (n=8), o mesmo número

de arestas ($n=17$) e nenhuma variável latente, as classes de equivalência contêm entre 2 a 6 DAGs.

Para modelos mais complexos, as classes de equivalência podem ser muito maiores. Por exemplo, se o esqueleto (o gráfico subjacente não direcionado) do DAG for completo (isto é, todos os vértices estão conectados), então a classe de equivalência contém exatamente $n!$ DAG, onde n é o número de vértices do gráfico (GILLISPIE & PERLMAN, 2002).

Diferente da descoberta de estruturas causais (*causal discovery*), a inferência causal requer que o pesquisador restrinja o espaço de modelos de acordo com o conhecimento a priori sobre o processo de geração dos dados. A escolha por um determinado modelo exclui não apenas os demais DAG da MEC, mas também outras possíveis configurações, incluindo variáveis latentes, variáveis tempo-dependentes, mecanismos de seleção, etc. Deste modo, uma tarefa fundamental é avaliar se o conhecimento disponível pode ser considerado suficiente para justificar essas escolhas.

Figura 12. Diferentes DAG e seus respectivos conjuntos de ajuste e a classe de modelos Markov equivalentes.



Legenda: na primeira coluna são apresentados 4 DAG (com diferentes esqueletos e imoralidades) contendo 8 vértices e 17 arestas e os respectivos conjuntos mínimos de ajuste (para o efeito total de E em D). Na segunda coluna, a classe de modelos markov equivalentes (MEC) de cada DAG é representada pelo cPDAG.

Fonte: CORTES, 2018.

Nos modelos contendo muitos vértices e arestas, a representação de vértices observados como vetores de variáveis pode facilitar a sua visualização. Esse tipo de

simplificação da estrutura gráfica por meio de “variáveis agregadas” foi um aspecto muito frequente nos DAG avaliados (n=59). Neste contexto, assume-se que os fatores agregados são afetados pelo exato conjunto de causas e, da mesma forma, todos os fatores causam o mesmo conjunto de efeitos. No entanto, em alguns casos, esses “vértices compostos” podem não ter um significado explícito. Por exemplo, em alguns artigos, os conjuntos de ajustes incluíam não somente os ancestrais de um vértice V, mas também as demais variáveis em V.

Para que uma ferramenta de inferência, como os diagramas causais, seja utilizada para uma comunicação eficaz é necessário consenso com relação às regras que prescrevem como os símbolos dessa linguagem devem ser utilizados e interpretados. Deste modo, a construção de modelos causais requer uma consideração adequada sobre como os componentes de uma teoria (conceitos, construtos) são operacionalizados e representados nos modelos (GRACE et al, 2012; BOLLEN & BAULDRY, 2011; SHAHAR & SHAHAR, 2010).

4.4.2 Relato de confundimento

A maioria dos artigos avaliados utilizou apenas o DAG (DAG – modelo total) na seleção dos confundidores para o modelo final (n=72). Uma das principais inconsistências na comunicação de confundimento observadas em pesquisas anteriores é a baixa frequência de relatórios que mencionam os critérios de seleção e inclusão dos confundidores nos modelos (GROENWOLD et al, 2008; POUWELS et al, 2015). Deste modo, a construção e apresentação de diagramas causais nos artigos representa um avanço na prática, garantindo que algumas das informações relevantes sobre confundimento estejam disponíveis para o leitor.

Em relação aos demais itens apresentados no STROBE, as frequências do relato entre os artigos com DAG foram semelhantes aos resultados observados em pesquisas anteriores (GROENWOLD et al, 2008; POUWELS et al, 2015). Alguns desses itens não estão diretamente relacionados a utilização do DAG causal, mas são aspectos importantes na comunicação de confundimento em estudos observacionais.

Mais da metade dos artigos avaliados (n=60) reconheceram a possibilidade de confundimento não observado. Os relatórios incluíram menos frequentemente as declarações qualitativas sobre a possível direção do viés (n=19) e o uso de análises de sensibilidade para a avaliação do potencial impacto do confundimento não observado (n=31). Em contraste, apenas 23 modelos gráficos incluíam ancestrais não observados da exposição e do desfecho e 4 apresentam causas comuns não observadas do mediador e do desfecho. Dependendo da estrutura causal, a presença de causas comuns não observadas pode ou não implicar em não identificação do parâmetro causal. Além disso, em alguns casos, a seleção do conjunto de ajuste pode ser feita por meio de outras estratégias de identificação (PEARL et al, 2014). Neste contexto, a consideração das fontes de incerteza sobre a estrutura gráfica e suas possíveis implicações para as conclusões derivadas do modelo, contribui para uma discussão mais transparente e unificada das limitações do estudo.

Poucas publicações (n=3) mencionaram algum tipo de informação sobre o número de conjuntos de ajuste derivados do DAG. No entanto, em alguns casos, o modelo gráfico pode implicar em mais de um conjunto suficiente (ou mínimo suficiente) de ajuste. Assumindo que a estrutura gráfica descreve corretamente o processo de geração dos dados, os conjuntos que satisfazem o critério da porta de trás implicam na mesma redução do viés de confundimento (PEAR & PAZ, 2014). Outra possível condição para a equivalência de confundimento é descrita em Pear & Paz (2014). Na prática, considerando aspectos de dimensionalidade, erros de mensuração e a medida de associação utilizada, é improvável que todos os conjuntos de ajuste derivados de um DAG sejam equivalentes. Considerações e alternativas ao uso de testes de colapsibilidade ou de independência para diagnósticos de modelos gráficos ou para seleção de covariáveis são apresentados em Greenland & Pearl (2011).

Para uma comunicação transparente de confundimento, recomenda-se que os autores incluam nos relatórios as informações sobre o tipo de estimativa de interesse (p.ex. efeito total, efeito indireto), o número e o tipo de conjunto de ajuste derivado do DAG e qualquer critério adicional utilizado na escolha dos conjuntos de ajuste.

Além disso, 20 publicações foram classificadas como possível ocorrência da falácia da tabela 2. Entre esses, apenas quatro artigos mencionaram possíveis diferenças na interpretação das estimativas. Como destacado em Westreich & Greenland (2013), o nome *table 2 fallacy* foi sugerido pelos autores, pois os resultados das análises de regressão múltipla são frequentemente apresentados na Tabela 2 dos artigos. A apresentação, em uma tabela, de todos os coeficientes estimados a partir de um único modelo, sugere, implicitamente, que todas as estimativas podem ser interpretadas da mesma forma. No entanto, uma das principais mensagens da teoria dos diagramas causais é que a interpretação das estimativas ajustadas pode variar, dependendo dos tipos de vértices condicionados ao longo dos caminhos no DAG. Deste modo, nem todos os coeficientes obtidos a partir de um modelo multivariado possuem a mesma interpretação causal, alguns coeficientes podem ser interpretados como um efeito total, alguns como efeito direto, e outros podem não ter nenhum significado causal.

Outro aspecto relacionado é a utilização de um único DAG para guiar as análises de diferentes exposições e/ou desfechos. Por definição, um DAG pode ser utilizado para responder diferentes perguntas causais a partir de intervenções do(.) em qualquer conjunto de variáveis no gráfico. No entanto, dependendo da pergunta de pesquisa, algumas partes da estrutura gráfica podem ter maior ou menor importância para a identificação do efeito causal.

A ferramenta de Web *Dagitty* apresenta uma seção específica contendo exemplos e exercícios relacionados ao tema da falácia da tabela 2 (disponível em: <http://www.dagitty.net/learn/graphs/table2-fallacy.html>).

5 RELATO DAS LIMITAÇÕES DO ESTUDO E MARCADORES DE INCERTEZA

5.1 Introdução

A comunicação transparente das fontes de incerteza de um estudo é fundamental para interpretação da validade dos resultados da pesquisa por parte da comunidade científica e pelo público geral e pode ter um papel fundamental no direcionamento de pesquisas futuras. No entanto, pesquisas anteriores sugerem que poucos trabalhos na literatura médica e epidemiológica discutem abertamente as possíveis fontes de incertezas que podem ter afetado os resultados e conclusões do estudo (IOANNIDIS, 2007; POCOCK et al, 2004; GROENWOLD et al, 2008; PAPATHANASIOU et al, 2010; SHARP et al, 2014; KARAHALIOS et al, 2012; POUWELS et al, 2015).

Nas últimas décadas, os diagramas causais têm sido apontados como uma das principais ferramentas que podem contribuir para a qualidade metodológica e do relato de estudos observacionais (PETERSEN & VAN DER LANN, 2014; PORTA, 2015; GLYMOUR, 2006). Os principais benefícios dos modelos causais gráficos são geralmente atribuídos à sua capacidade de comunicar de forma transparente o conhecimento sobre o processo de geração de dados, as limitações desse conhecimento e avaliar o que pode ser inferido, dado essas premissas. Por outro lado, também há uma preocupação de que, dependendo da maneira como são utilizadas, essas ferramentas podem colaborar para inferências excessivamente confiantes (GREENLAND, 2017a; DAWID, 2010; GREENLAND, 2010).

Neste trabalho, é apresentada uma análise comparativa preliminar de três aspectos da comunicação de fontes de incerteza em artigos com e sem diagramas causais. Foram investigados o número de limitações do estudo reconhecidos pelos autores, o tipo de avaliação das limitações e a frequência de marcadores linguísticos de incerteza. As análises apresentadas neste capítulo são utilizadas para discutir algumas dificuldades metodológicas relacionadas a formulação da pergunta de pesquisa sob uma perspectiva causal e a construção de diagramas causais.

5.2 Métodos

5.2.1 Seleção dos estudos

As estratégias de busca e seleção dos estudos com a utilização de diagramas causais foram descritas na seção 4.1. De forma geral, após a execução das estratégias de busca e avaliação dos registros, foram incluídos 1034 artigos com o relato da utilização de diagramas causais. Posteriormente, foram identificadas as publicações que apresentaram a estrutura gráfica no artigo ou no material suplementar (N=430). Desse total de artigos foi selecionada uma amostra aleatória de N=100.

Para cada artigo da amostra, foi selecionado aleatoriamente um artigo de comparação publicado na mesma revista e período. Primeiro, as informações sobre a revista, ano e número (*issue*) dos artigos com DAG foram exportadas para um banco de dados por meio do programa EndNote. No ambiente R, essas informações foram combinadas em uma única chave de busca e executadas por meio da biblioteca RisMed.

Com o objetivo de minimizar a necessidade de realização de novas buscas, foram considerados inicialmente 2 artigos de comparação. Os critérios de inclusão e exclusão dos controles foram os mesmos do grupo com DAG, exceto no aspecto da utilização de diagramas causais. Em todos os casos, foram escolhidos como controle o primeiro artigo que se adequava aos critérios de elegibilidade. Todos os artigos com DAG e os artigos de comparação (Sem DAG) foram analisados quanto ao relato das limitações do estudo e a frequência de marcadores de incerteza (*hedges*).

5.2.2 Coleta de dados

5.2.2.1 Relato das limitações do estudo

Os itens sobre o relato das limitações do estudo foram elaborados tendo como base o item 19 do STROBE e publicações anteriores sobre o tema (TER RIE et al, 2013). A versão final do questionário de extração de dados é apresentada no APÊNDICE C.

Foi avaliado o número de limitações do estudo ou vieses em potencial reconhecido pelos autores, em qualquer parte do artigo, e os tipos de limitações relatadas. Além disso, foram extraídas as informações sobre o tipo de avaliação de cada limitação: a) se os autores discutiram qualitativamente a possível magnitude e direção do viés e b) se os autores relataram análises de sensibilidade para avaliar o possível impacto da limitação e o tipo de análise relatada.

Os tipos de avaliações das limitações do estudo foram classificados como: a) **nenhuma avaliação**, quando os autores apenas reconheceram as limitações (sim ou não), b) **avaliações qualitativas**: quando os autores discutiram qualitativamente pelo menos uma limitação do estudo (sim ou não) e c) **avaliações quantitativas**: quando os autores relataram pelo menos uma análise de sensibilidade (sim ou não).

As avaliações quantitativas foram classificadas como **a) técnicas não probabilísticas** (fórmulas de viés): que permitem calcular algebricamente a magnitude do viés necessária para explicar a associação observada; **b) técnicas de modelagem de viés**: foram incluídas nessa categoria as técnicas que utilizam uma distribuição dos parâmetros do viés para reajustar as medidas de efeito e precisão (*Probabilistic bias analysis, Bayesian sensitivity analysis, Monte Carlo sensitivity analysis*), técnicas de imputação múltipla de dados e técnicas baseadas em Ponderação pela Probabilidade Inversa (IPCW, IPTW); **d) outros tipos de análise de sensibilidade**: em geral, estratégias envolvendo a inclusão ou exclusão de variáveis no modelo, restrição da amostra e a redefinição de variáveis.

5.2.2.2 Marcadores linguísticos de incerteza

A amostra de artigos com diagramas causais e os artigos de comparação foram analisados quanto a frequência de marcadores de incertezas (*hedges*).

Existem vários fenômenos linguísticos que são referidos como incerteza na literatura (CLEMEN, 1997). Neste trabalho, foi avaliada apenas a incerteza semântica identificada por meio de pistas lexicais. Para a seleção dos itens lexicais foram considerados os itens apresentados em Hyland (2004), referente a um estudo de 500.000 corpus de artigos de pesquisa acadêmica de oito disciplinas.

O pré-processamento do corpus incluiu a conversão dos arquivos em PDF para o formato texto (utilizando a ferramenta *pdftotext* – disponível em: <http://www.xpdfreader.com>) e a extração manual da seção de discussão, considerando o formato (layout) dos artigos. Após o pré-processamento do corpus, todos os 96 itens foram avaliados separadamente quanto a frequência e ambiguidade. Os itens que não funcionavam como *hedges* em muitos contextos foram excluídos da lista final. Além disso, foram adicionadas a lista original as possíveis variações dos verbos. A lista final contendo 90 itens (e suas variações) é apresentada no APÊNDICE D. Os dados foram analisados no programa de acesso aberto, AnCont Versão 3.5.5, disponível: <http://www.laurenceanthony.net/>)

5.2.2.3 Características gerais das publicações e dos coautores

Foram extraídas as informações sobre o desenho de estudo, objetivos primários e secundários do estudo (avaliação do efeito total, análise de mediação, modificação de efeito e outros) e o tamanho da amostra.

O gênero foi atribuído aos autores com base no primeiro nome por meio da ferramenta de web *Genni* (disponível em <http://abel.lis.illinois.edu>). Essa ferramenta utiliza um modelo de predição baseado em dados públicos sobre a frequência de um nome entre homens e mulheres ao longo do tempo, e inclui informações de outros países além dos Estados Unidos (Smith et al, 2013). Em comparação com o banco de dados do US Social Security Administration (SSA), o modelo apresentou erro de classificação de 2,25% dos nomes incluídos nos bancos de dados. As buscas na ferramenta *Genni* foram realizadas por meio do número de identificação do artigo no pubmed (PMID). Na predição do gênero, os nomes dos autores são classificados como masculinos, femininos, neutros ou desconhecidos. Para os casos neutros e

desconhecidos foram realizadas diferentes buscas, incluindo a ferramenta de autores da base de dados Scopus (para determinação do nome completo e dados de afiliação) e pesquisas nos sites da universidade de afiliação. A ferramenta *Genni* também permite a predição da etnia com base no nome dos autores (TORVIK & AGARWAL, 2016). Os países de origem foram classificados de acordo com a língua oficial (inglesa e não inglesa).

5.2.3 Análises de dados

Foram considerados três desfechos principais, a frequência de termos de *hedges*, o número de limitações do estudo reconhecidos pelos autores e o tipo de avaliação das limitações. Considerando que os tamanhos das seções de discussão podem ser diferentes entre os grupos, as informações sobre os termos *hedges* são apresentadas como a frequência média normalizada por 100 palavras (GRIES, 2008). A frequência normalizada de termos *hedges* foi calculada por meio da seguinte fórmula: (frequência de termos *hedges*/número de *tokens* x 100). A normalização foi realizada considerando todas as palavras (*tokens*).

O número de limitações reconhecidas nos artigos e a frequência média de termos *hedges* por 100 palavras foram comparados por meio de modelos de regressão linear. Para as variáveis categóricas foram estimadas razões de prevalência por meio de modelos de regressão log-binomial. As análises estatísticas foram conduzidas no ambiente R - versão 3.0.2 (R Foundation for Statistical Computing, www.r-project.org, Vienna, Austria). A versão reduzida dos scripts e as bibliotecas utilizadas são apresentadas no APÊNDICE E.

5.3 Resultados

A tabela 5 apresenta as características gerais dos estudos. 57% dos artigos avaliados eram de estudos longitudinais. A maioria das publicações realizou análises de efeito total (n=148, 74%). A frequência de análises de mediação foi maior entre

os artigos com DAG (16% vs. 4%). A mediana do tamanho da amostra nos artigos com diagramas causais foi de 2.550 (IQR 1025-13.667) e 4.289 (IQR 900-13.782) nos artigos sem DAG.

A frequência média de termos *hedges* por 100 palavras foi semelhante nos dois corpora (DAG 1,8% vs. 1,7% nos artigos sem DAG, diferença de 0,1 IC 95% 0,05 – 0,29). (Tabela 6). A média de limitações reconhecidas pelos autores foi de 4,2 nas publicações com diagramas causais e 3,6 nos artigos sem DAG, com uma diferença média de 0,4 (IC 95% de -0,01 – 0,86).

Alem disso, 78 publicações (39%) apenas reconheceram as limitações do estudo, enquanto 95 avaliaram qualitativamente pelo menos uma limitação (46%). As limitações reconhecidas pelos autores referiam-se principalmente aos aspectos de erros de mensuração (n= 158, 79%) (Gráfico 3).

A possibilidade de confundimento não observado foi reconhecida em 60% (IC 95% 50- 69) dos artigos com DAG e em 42% (IC 95% 42 - 52) dos artigos sem DAG. O relato de avaliações quantitativas das limitações do estudo foi mais frequente entre os artigos com diagramas causais (RP 2,1; IC 95% 1,62 - 3,78). No entanto, a utilização de técnicas de modelagem de viés foi menos frequente nos dois grupos (8% e 5%)

Tabela 6. Características gerais dos artigos e coautores.

Item	Com DAG (N=100)	Sem DAG (N=100)
	%	%
Desenho de estudo		
Longitudinal	60	54
Transversal	30	31
Caso controle	10	15
Objetivo principal do estudo		
Análise do efeito total	68	80
Análise de mediação	16	4
Análises de fatores de risco	11	8
Outros	5	8
Tamanho da amostra[§]	2.550 (1025-13.667)	5275 (775-16.762)
Número de coautores[§]	7 (4,7-9,0)	7 (5,0-8,2)
Gênero do primeiro autor *		
Feminino	63	59
Masculino	36	40
<i>Incerto</i>	1	1
Língua oficial do país de origem*		
Inglesa	33	22
Não inglesa	66	77
<i>Incerto</i>	1	1
Número de palavras[§]	1.450 (1.236-1.730)	1.384 (1.050-1.754)
Número de termos hedges[§]	26 (18-35)	22 (16-29)

Nota: *Informações referentes ao primeiro autor do manuscrito.

[§]As variáveis numéricas são apresentadas por medianas e intervalos interquartis.

Fonte: CORTES, 2018.

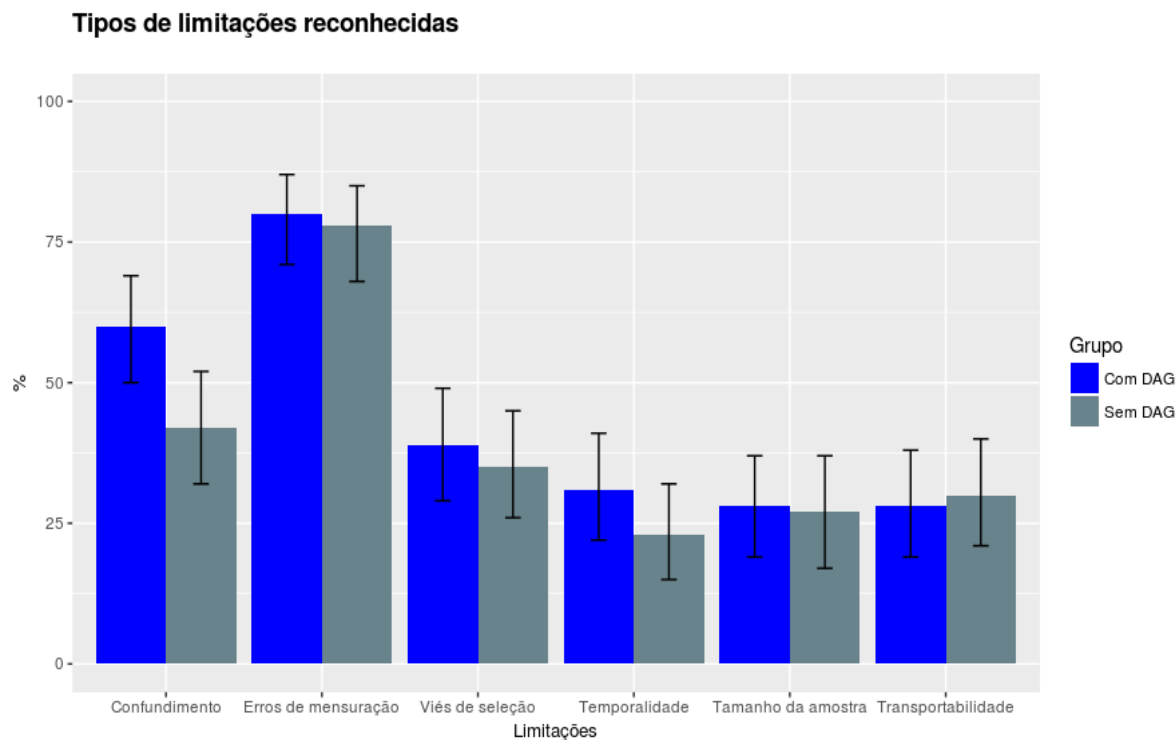
Tabela 7. Frequência de termos hedges e reconhecimento das limitações do estudo em artigos com e sem diagramas causais.

Item	Com DAG (N=100)	Sem DAG (N=100)	Diferença (IC 95%)
Frequência média de termos hedges*	1,84	1,72	0,12 (-0,05 – 0,29)
Número de limitações reconhecidas	4,18	3,75	0,43 (-0,01 – 0,86)
Tipo de avaliação das limitações			RP (IC 95%)
Apenas reconheceu as limitações	31	47	0,71 (0,54 – 0,94)
Avaliou qualitativamente	50	45	1,09 (0,82 – 1,44)
Avaliou quantitativamente	52	21	2,09 (1,62 – 3,78)
<i>Modelagem de viés</i>	8	5	
<i>Fórmula de viés</i>	4	1	

Nota: * Frequência média de termos hedges a cada 100 palavras na seção de discussão.

Fonte: CORTES, 2018.

Gráfico 3. Tipo de limitações do estudo e fontes de viés reconhecidas pelos autores em artigos com e sem diagramas causais.



5.4 Discussão

O número de limitações do estudo reconhecidas pelos autores e o nível de atenuação da linguagem (frequência de marcadores de incerteza) foram semelhantes nas amostras de artigos com e sem diagramas causais. A utilização de técnicas de análise de sensibilidade foi mais frequente entre os artigos com DAG causal. No entanto, principalmente para as variáveis numéricas, as estimativas são muito imprecisas para serem informativas.

A pesquisa sobre o reconhecimento das limitações do estudo ainda é escassa. Anteriormente, em uma revisão de 300 artigos publicados em revistas biomédicas, Rie et al (2013) observaram que 27% (n=73) das publicações não reconheceram qualquer limitação do estudo (27%) e 73% publicações reconheceram uma mediana de 3 limitações (1-8) na seção de discussão. Semelhante ao presente estudo, as fontes de erros sistemáticos mais relatados foram aquelas decorrentes de erros de mensuração e de confundimento (TER RIE et al, 2013).

As análises apresentadas neste capítulo são consideradas exploratórias; além da limitação do tamanho da amostra, são discutidas nas próximas seções outras dificuldades metodológicas relacionadas a interpretação pergunta de pesquisa sob uma perspectiva causal e possíveis estruturas causais para o tema.

5.4.1 Interpretações da pergunta de pesquisa

Neste trabalho, a “exposição” de interesse foi restrita a apresentação no artigo de um diagrama causal na forma de um DAG (versus o não relato da utilização dessas ferramentas). Uma possível formulação da pergunta de pesquisa sob uma perspectiva causal seria: o que aconteceria com o nível de autocrítica (relato das limitações) e o nível de atenuação da linguagem nos relatórios de estudos observacionais se forçarmos os autores a apresentarem no artigo as suas hipóteses causais na forma de um DAG causal? Uma resposta comum para essa pergunta é, “depende de como essas ferramentas serão utilizadas”. Os diagramas causais podem ser aplicados em diferentes fases da pesquisa, desde o planejamento do estudo, na execução das análises e no decorrer da redação do artigo. Podemos supor que a utilização dessas ferramentas em fases mais precoces da pesquisa pode contribuir para uma melhor especificação dos modelos estatísticos e avaliação das fontes de viés. Do mesmo modo, como um diagrama causal é “utilizado” pode depender de outros fatores como os objetivos das análises, o conhecimento dos autores sobre o tema de pesquisa e sobre os pressupostos e regras da teoria dos diagramas causais.

Nota-se que a descrição apresentada acima é semelhante aos exemplos de múltiplas versões da exposição apresentados por alguns autores (COLE & FRANGAKIS, 2009; VANDERWEELE, 2009; VANDERWEELE & HERNÁN, 2013) - o que poderia implicar em violação da “hipótese” de consistência. Nesse contexto, a não consistência dos desfechos potenciais ocorre quando há múltiplas formas de se atribuir o mesmo valor ($X=1$) à uma variável de exposição X , e algumas dessas formas podem ter um impacto diferenciado sobre o desfecho. Para os autores, uma maneira de lidar com as dificuldades geradas por múltiplas versões da exposição

seria redefinir cada versão como uma exposição distinta e restringir as inferências de acordo com o tipo de versão avaliada (VANDERWEELE & HERNÁN, 2013). Além disso, dependendo das relações entre a exposição composta e a versão da exposição avaliada (e da estimativa de interesse), teríamos que considerar, não apenas, as causas comuns entre a versão da exposição e o desfecho, mas também os fatores de confusão que influenciam as versões do tratamento (*treatment-version confounder*).

Como destacado em Cole & Frangakis (2009), uma das motivações para a consideração dessas questões como um requisito para a inferência causal, refere-se à noção de que algumas perguntas causais são mais significativas para a formulação de decisões políticas do que outras. Na prática, a especificação de intervenções exige, muitas vezes, outras informações além das estimativas de efeito de intervenções ideais. Em outras palavras, saber o quão eficaz é um tratamento não nos diz o quão eficaz é qualquer política que se destina a administrar esse tratamento na prática (PEARL, 2014b).

No contexto da teoria dos diagramas causais, todas as perguntas causais definidas como intervenções *do(.)* em quaisquer variáveis no gráfico, são consideradas bem definidas (isto é, consistentes) (PEARL, 2010b; 2014b). Além disso, a especificação de intervenções compostas, do tipo P ($Y \mid do(x), do(w), do(z)$) (p.ex., uma política editorial referente aos DAG em conjunto com um programa de ensino sobre o tema, etc.), também são bem definidas, desde que essas sejam observadas e descritas no modelo (PEARL, 2010b; 2014b).

Há dois pontos principais a serem destacados. Primeiro, apesar de ser relativamente fácil de se conceber uma intervenção ideal referente aos DAG (p.ex., na forma de uma política editorial), podemos destacar outras perguntas possivelmente mais relevantes. Por exemplo, que fatores podem estar relacionados à utilização dos DAG ou à construção de modelos “mais realistas” (colaboração em pesquisa, áreas de pesquisa, características dos autores). Segundo, por questões práticas, a “exposição” foi definida como a apresentação do DAG no artigo versus o não relato da utilização dessas ferramentas. No entanto, uma avaliação mais apropriada consideraria também aqueles artigos que relataram a utilização dos

diagramas causais, mas que não apresentaram a estrutura (isto é, outra possível versão da intervenção).

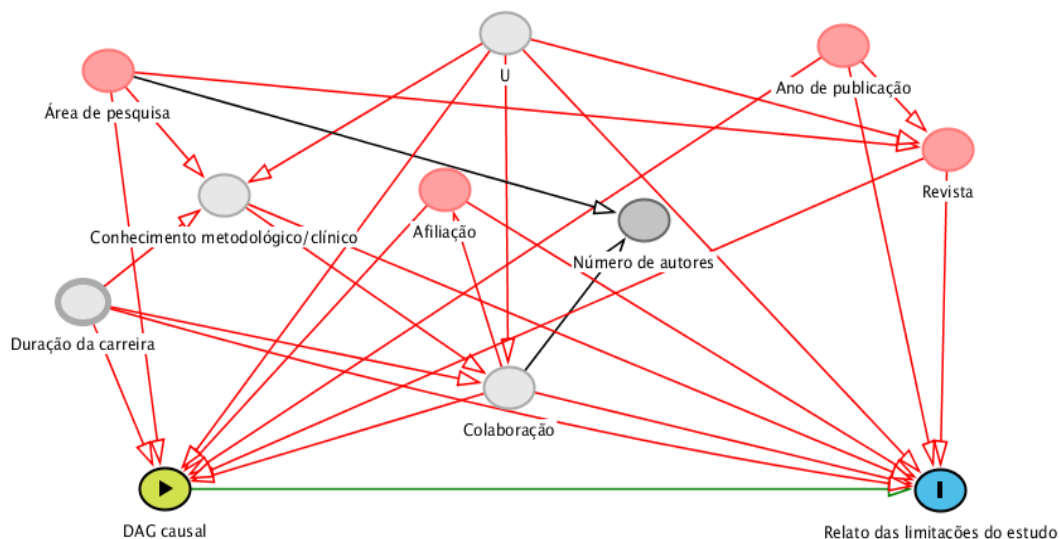
5.4.2 Estruturas causais

Como hipótese geral, parece razoável supor que as atitudes em relação à comunicação de fontes de incertezas em artigos científicos dependem de fatores ligados à sua percepção e aos incentivos e restrições à sua comunicação. De forma geral, os diagramas causais são ferramentas que permitem comunicar de forma transparente o conhecimento sobre o processo de geração de dados, as limitações desse conhecimento, e avaliar o que pode ser inferido, dado essas premissas.

A primeira dificuldade para a construção do modelo causal é que há poucas informações sobre os fatores específicos que podem influenciar o uso (e a apresentação) dos DAG e os desfechos avaliados. No entanto, tendo como base a literatura sobre comunicação de incertezas e sobre a qualidade dos relatórios de pesquisa, foram considerados alguns fatores gerais potencialmente relacionados à comunicação transparente das hipóteses do estudo e de suas limitações (Gráfico 3).

Por exemplo, assume-se que alguns fatores individuais, como aspectos cognitivos gerais (DUNCAN, 1972; TVERSKY & KAHNEMAN, 1979), traços de personalidades (HODSON, 1999), o conhecimento específico do domínio e crenças sobre o conhecimento (BUEHL, 2001), podem afetar as atitudes e expectativas sobre como o trabalho deve ser conduzido e comunicado. Alguns desses fatores são apresentados separadamente no DAG (Gráfico 3), como vértices latentes, outros são representados pelo vértice latente U. Além disso, o DAG assume que as potenciais influências do processo de revisão por pares e aspectos da política editorial, como a duração do processo de revisão e as características dos revisores/editores, podem ser mensuradas pela revista de publicação (e o período).

Gráfico 4. Diagrama causal representando as hipóteses de sobre a relação entre a apresentação de diagramas causais em estudos observacionais e o relato das limitações do estudo.



Legenda: Os vértices em cinza indicam variáveis não observadas (latentes).

Fonte: CORTES, 2018.

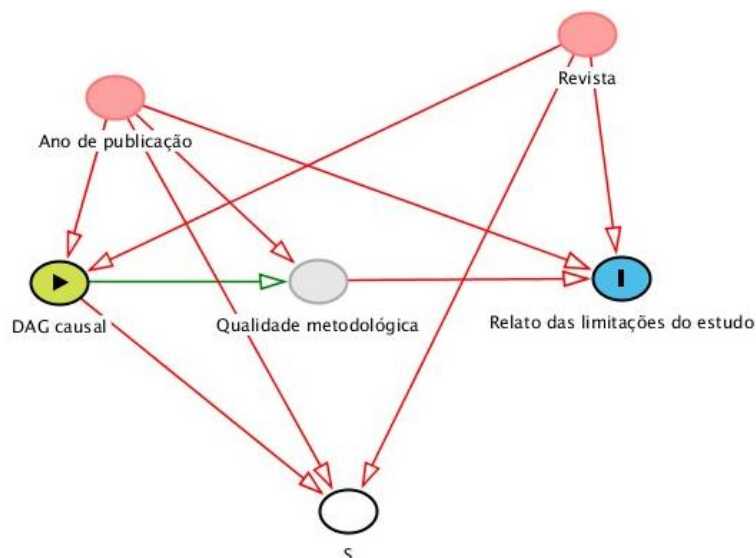
A colaboração em pesquisa pode contribuir para a troca de ideias e recursos, a aprendizagem de novas habilidades e a resolução de problemas complexos (BUKVOVA, 2010). A colaboração em pesquisa pode assumir diferentes formas e envolver outros aspectos além da afiliação dos autores e da coautoria (KATZ & MARTIN, 1997). As decisões sobre a sua operacionalização e representação dependerá das hipóteses sobre como os aspectos da colaboração podem influenciar a comunicação transparente das hipóteses causais por meio dos DAG e o relato de fontes de incerteza. Em outras palavras, um aspecto importante é o significado dado a cada vértice no gráfico e as hipóteses sobre como cada vértice se relaciona com os processos no sistema de interesse.

O processo de concepção, construção e publicação de um artigo científico envolve diferentes etapas e interações entre diferentes atores, incluindo os coautores, editores, revisores, editores técnicos e de redação, dentre outros. Essa descrição sugere que os modelos considerados podem não ser representações adequadas para o tema.

Um aspecto relacionado refere-se à seleção do grupo de comparação. A decisão do pareamento por revista e período teve como base a estrutura do Gráfico 4. Neste caso, o objetivo é semelhante ao pareamento em estudos de coorte e, na ausência de outros fatores de confusão, o ajuste por essas variáveis pode ser desnecessário (MANSOURNIA et al, 2013). No entanto, a independência (*unfaithfulness*) criada pelo pareamento, geralmente, será quebrada após o ajuste para outras covariáveis (SJÖLANDER & GREENLAND, 2013). Deste modo, considerando o Gráfico 3, um conjunto de ajuste parcial deveria incluir as variáveis do pareamento e os demais confundidores observados: {Ano de publicação, Revista, Afiliação, Área de pesquisa}.

A razão de prevalência para o relato de avaliações quantitativas das limitações do estudo foi de 2,09 (IC 95%, 1,62 – 3,78). Por meio da fórmula apresentada em VanderWeele & Ding (2017) podemos calcular a força máxima de um confundidor não mensurado que poderia explicar completamente as estimativas observadas. A RP observada de 2,09, pode ser explicada por um confundidor associado a exposição e ao desfecho com uma RP de 3,6, para ambos. Destaca-se que o valor de uma fórmula de viés deve ser interpretado com cautela; dado que o valor calculado é condicionado às covariáveis ajustadas e, obviamente, outros ajustes podem implicar em valores inferiores ou superiores (p.ex., na presença de confundimento negativo) da fórmula viés.

Gráfico 5. Gráfico acíclico direcionado simplificado representando as hipóteses de seleção.

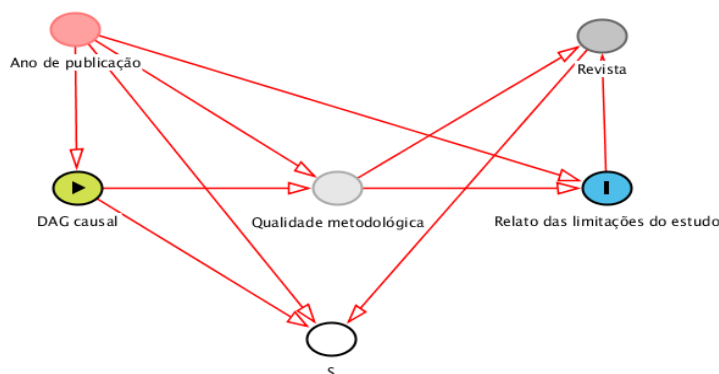


Legenda: o condicionamento pelo vértice S indica a seleção no estudo.

Fonte: CORTES, 2018.

Por outro lado, as decisões editoriais sobre a publicação de um artigo podem ser pautadas em vários aspectos, incluindo as características do relatório. Como representado no gráfico 5, se a fonte de publicação do artigo é influenciada pelas características dos relatórios, o pareamento pela revista de publicação pode implicar em viés de colisão. De um modo geral, um grande viés de colisão exige fortes associações entre o colisor e a exposição e o desfecho. Além da relação positiva entre a exposição e a seleção no estudo, parece plausível, principalmente nos casos dos marcadores de incerteza e do tipo de avaliação das limitações, que também tenha uma relação positiva entre os desfechos e a seleção.

Gráfico 6. Gráfico acíclico direcionado simplificado representando hipóteses alternativas de seleção.



Legenda: o condicionamento pelo vértice S indica a seleção no estudo.

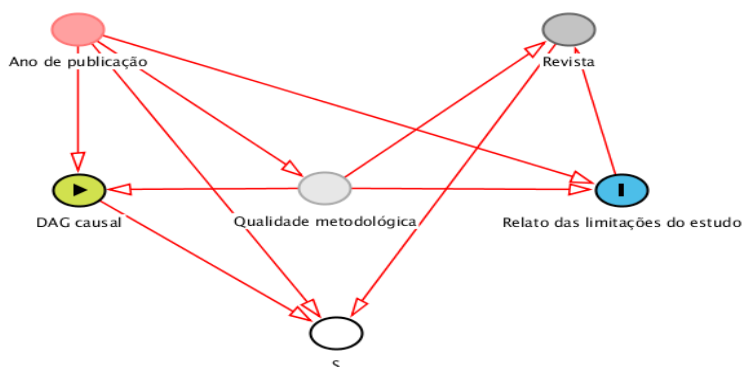
Fonte: CORTES, 2018.

Algumas condições para avaliação da direção do viés de colisão são apresentadas em Jiang & Ding (2016) e Nguyen et al (2017). No caso de variáveis dicotômicas e na escala de Razão de chances, Jiang & Ding (2017) demonstraram que, se a exposição e o desfecho afetam a seleção na mesma direção e têm interação não positiva em S, então espera-se que a estimativa na população selecionada corresponda a um limite inferior para o parâmetro. De forma geral, (e no mesmo contexto) se E e D afetam S na mesma direção e eles não têm interação em S, então o condicionamento em S introduzirá uma associação negativa entre E e D. Por outro lado, se a utilização de DAG causal não afeta o tipo de avaliação das limitações do estudo (isto é, não há uma seta entre E e D), então o sinal do viés de colisão em um nível específico de S não é dependente da escala de efeito.

No entanto, os gráficos 4 e 5 parecem igualmente plausíveis, dado que grande parte dos elementos discutidos anteriormente não são atributos estáticos. Nesses casos, seria necessário estender o modelo e as observações ao longo do tempo. Outra possível interpretação é que a utilização de DAG causais ou apresentação do modelo gráfico no artigo seja influenciada pelo nível de incerteza dos autores sobre o tema de pesquisa ou à qualidade metodológica do estudo

(Gráfico 6). Podemos interpretar a qualidade metodológica do estudo, por exemplo, como a qualidade e quantidade dos dados disponíveis.

Gráfico 7. Gráfico acíclico direcionado simplificado representado as hipóteses alternativas de seleção e confundimento.

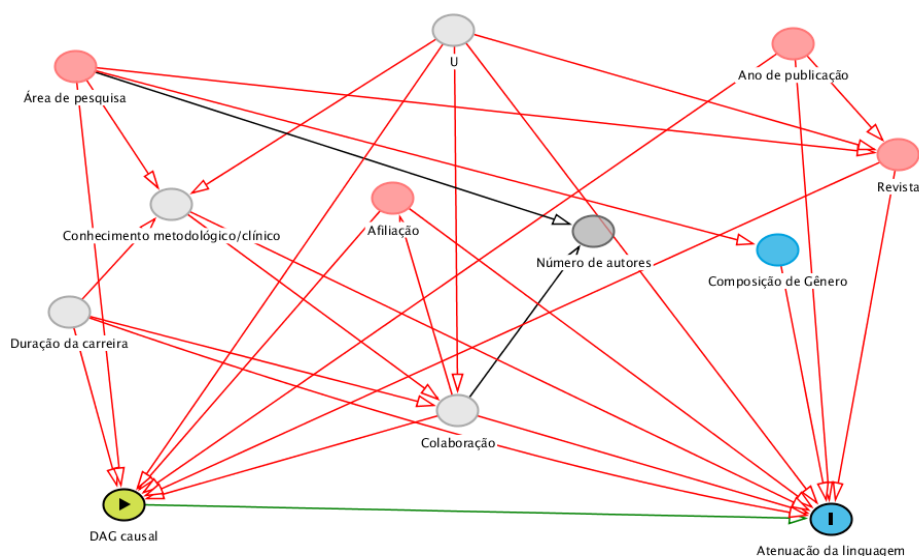


Legenda: o condicionamento pelo vértice S indica a seleção no estudo.

Fonte: CORTES, 2018.

Para o desfecho relacionado à frequência de termos hedges (nível de atenuação da linguagem) foram consideradas outras potenciais relações apresentadas no Gráfico 7. Alguns estudos sugerem que a expressão de incerteza no discurso escrito pode variar segundo a língua dos autores (MIRZAPOUR & MAHAND, 2012; FARROKHI et al, 2008; KOURILOVA, 1995). Os autores cuja primeira língua não é o inglês, podem transferir os padrões discursivos típicos de sua língua e cultura de origem, podendo levar a falhas na aplicação de estratégias retóricas, incluindo os recursos de atenuação (HINKEL, 1999). No DAG, o país de afiliação do primeiro autor é visto como uma causa, não apenas, da habilidade de escrita em inglês, mas também de outros fatores relacionados à cultura acadêmica e outras condições socioculturais (modelo de ensino, incentivos). Deste modo, uma possível medida mais apropriada seria determinar o país de afiliação de todos os coautores com base nos dados de afiliação ao longo de suas carreiras (FANELLI et al, 2015).

Gráfico 8. Diagrama causal representado as hipóteses sobre a relação entre a apresentação de diagramas causais em estudos observacionais e o nível de atenuação da linguagem.



Legenda: Os vértices em cinza indicam variáveis não observadas (latentes).

Fonte: CORTES, 2018.

Outra questão refere-se à tradução das evidências disponíveis na literatura em hipóteses causais codificadas no DAG. Por exemplo, alguns autores sugerem diferenças de gênero na linguagem tentativa falada (LEAPER & ROBNETT, 2011). No entanto, há poucas evidências disponíveis sobre o papel do gênero na expressão de incerteza no discurso escrito acadêmico. Essa incerteza é representada no DAG (Gráfico 7) pela seta entre o vértice composição de gênero e a atenuação da linguagem. No entanto, a exclusão da seta entre a composição de gênero do artigo e a apresentação de diagramas causais (ou seus ancestrais) pode não ter justificção empírica. Além disso, como destacado em Greenland (2010), uma questão central é como podemos justificar a remoção de setas no gráfico, dado que as evidências para tal remoção serão baseadas, muitas vezes, em conclusões sobre ausência de associação entre as variáveis; que por sua vez dependem das demais hipóteses

5.4.3 Limitações

O número reduzido de artigos avaliados e ausência de dados adequados para avaliação da confiabilidade das medidas são algumas das limitações deste trabalho. Como destacado em Murad & Wang (2017) a quantificação da concordância entre os revisores é ponto mais crítico em um estudo meta epidemiológico do que em uma revisão sistemática clínica, tendo em vista os julgamentos (mais subjetivos) feitos na classificação dos dados em um estudo meta epidemiológico.

Estudos futuros podem considerar a implementação de abordagens de automatização do processo de revisão. Nos últimos anos, têm sido apresentados projetos de automatização ou simplificação de revisões sistemáticas (JONNALAGADDA et al, 2015; HOWARD et al 2016; MARSHALL et al, 2016; TSAFNAT et al, 2014), incluindo a extração de termos, como o p-valor (CHAVALARIAS et al, 2016) e outras informações metodológicas (p.ex, KIRITCHENKO, et al 2010). No entanto, o estado atual da pesquisa de extração automática de dados ainda é exploratório (JONNALAGADDA et al, 2015) e grande parte dos itens avaliados neste trabalho não estão incluídos nas ferramentas atualmente disponíveis.

Uma das limitações na avaliação da ocorrência de termos hedges foi a utilização de uma lista reduzida de pistas lexicais, que pode ser expandida por meio de abordagens semelhantes (KILICOGU & BERGLER, 2008). Algumas alternativas incluem as estratégias baseadas em regras ou abordagens de aprendizagem supervisionada (ver, FARKAS et al, 2010).

A ordem do texto é importante na identificação da função de hedges e na análise descritiva do corpus. Nos principais programas de extração de texto de arquivos de PDF (p.ex., Adobe Acrobat e PDF2Text), o fluxo do texto pode ser desordenado devido a presença de legendas, tabelas, figuras, etc. Além disso, dado à variedade de formatos dos artigos entre revistas (e entre autores), a identificação e extração automática de seções específicas dos artigos não é tão trivial. Neste trabalho, a extração da seção de discussão dos artigos foi realizada manualmente. Uma ferramenta de livre acesso para extração do texto (completo)

em PDF com reconhecimento do layout é apresentada em Ramakrishnan et al (2017).

CONCLUSÃO

O processo de modelagem causal pode ser caracterizado por, pelo menos, três escolhas: a escolha das variáveis que incluímos nos modelos, do conjunto de equações (ou setas) que descrevem as relações entre elas, e quais variáveis são determinadas como fatores endógenos ou exógenos. Além disso, essas escolhas geralmente ocorrem em um contexto de demandas contrastantes, envolvendo tensões entre as nossas limitações práticas e cognitivas, os objetivos da pesquisa e a complexidade dos fenômenos estudados.

Os potenciais benefícios da utilização de ferramentas de inferência, como os diagramas causais dependem do entendimento correto de suas regras e de suas potenciais limitações (p.ex, AALEN et, 2016; GREENLAND & MANSOURNIA, 2015). Como destacado em Greenland (2017a), um aspecto central é como ensinar o uso correto de métodos formais e como aplicá-los sem gerar excessos de confiança e interpretações errôneas que arruinaram tantas práticas estatísticas.

Essas questões reforçam, não somente, a necessidade de expansão do vocabulário metodológico para incluir outros tipos de vieses e processos cognitivos, mas também a formulação de programas de pesquisa interdisciplinares direcionados ao estudo da prática de pesquisa (GREENLAND, 2017B; IOANNIDIS, 2018).

Por fim, destaca-se que a avaliação da adequação dos modelos causais exige a consideração dos aspectos teóricos e metodológicos de cada domínio específico e diferentes pesquisadores podem construir diferentes estruturas causais. No entanto, espera-se que esse trabalho possa contribuir para a conscientização de alguns aspectos importantes relacionados a construção, análise e apresentação dos DAG causais.

REFERÊNCIAS

- AALEN, Odd Olai et al. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. **Statistical methods in medical research**, v. 25, n. 5, p. 2294-2314, 2016.
- ALLAHYARI, Mehdi et al. A brief survey of text mining: Classification, clustering and extraction techniques. **arXiv preprint arXiv:1707.02919**, 2017.
- AUCHINCLOSS, Amy H.; ROUX, Ana V. Diez. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. **American journal of epidemiology**, v. 168, n. 1, p. 1-8, 2008.
- BAMMER, Gabriele; SMITHSON, Michael (Ed.). **Uncertainty and risk: multidisciplinary perspectives**. Routledge, 2008.
- BAREINBOIM, Elias; PEARL, Judea. A general algorithm for deciding transportability of experimental results. **Journal of causal Inference**, v. 1, n. 1, p. 107-134, 2013.
- BAREINBOIM, Elias; PEARL, Judea. Causal inference and the data-fusion problem. **Proceedings of the National Academy of Sciences**, v. 113, n. 27, p. 7345-7352, 2016.
- BAREINBOIM, Elias; PEARL, Judea. Controlling selection bias in causal inference. In: **Artificial Intelligence and Statistics**. 2012. p. 100-108.
- BAREINBOIM, Elias; PEARL, Judea. Transportability of causal effects: completeness results. In: **Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence**. AAAI Press, 2012. p. 698-704.
- BARRETO, Mauricio Lima et al. Differences between h-index measures from different bibliographic sources and search engines. **Revista de saude publica**, v. 47, n. 2, p. 231-238, 2013.
- BASTUJI-GARIN, Sylvie et al. Impact of STROBE statement publication on quality of observational study reporting: interrupted time series versus before-after analysis. **PloS One**, v. 8, n. 8, p. e64733, 2013.
- BEKELMAN, Justin E.; LI, Yan; GROSS, Cary P. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. **Jama**, v. 289, n. 4, p. 454-465, 2003.
- BERKSON, Joseph. Limitations of the application of fourfold table analysis to hospital data. **Biometrics Bulletin**, v. 2, n. 3, p. 47-53, 1946.
- BIERI, James. Cognitive complexity-simplicity and predictive behavior. **The Journal of Abnormal and Social Psychology**, v. 51, n. 2, p. 263, 1955.

BITTI, Pio E. Ricci et al. Expression and communication of doubt/uncertainty through facial expression. **Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education**, v. 9, n. 1, p. 159-177, 2014.

BLANKENSHIP, Kevin L.; HOLTGRAVES, Thomas. The role of different markers of linguistic powerlessness in persuasion. **Journal of Language and Social Psychology**, v. 24, n. 1, p. 3-24, 2005.

BOLGER, Donald J. et al. The role and sources of individual differences in critical-analytic thinking: a capsule overview. **Educational Psychology Review**, v. 26, n. 4, p. 495-518, 2014.

BOLLEN, Kenneth A.; BAULDRY, Shawn. Three Cs in measurement models: Causal indicators, composite indicators, and covariates. **Psychological methods**, v. 16, n. 3, p. 265, 2011.

BRENNER, Hermann; BLETTNER, Maria. Misclassification Bias Arising from Random Error in Exposure Measurement: Implications for Dual Measurement Strategies. **American journal of epidemiology**, v. 138, n. 6, p. 453-461, 1993.

BREITLING, Lutz Philipp. dagR: a suite of R functions for directed acyclic graphs. **Epidemiology**, v. 21, n. 4, p. 586-587, 2010.

BRIGGS, David J.; SABEL, Clive E.; LEE, Kayoung. Uncertainty in epidemiology and health risk and impact assessment. **Environmental geochemistry and health**, v. 31, n. 2, p. 189-203, 2009.

BROADBENT, Alex. Causation and prediction in epidemiology: A guide to the "Methodological Revolution". **Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences**, v. 54, p. 72-80, 2015.

BROSS, Irwin. Misclassification in 2x2 tables. **Biometrics**, v. 10, n. 4, p. 478-486, 1954.

BRUCE, Rachel et al. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. **BMC medicine**, v. 14, n. 1, p. 85, 2016.

BUEHL, Michelle M.; ALEXANDER, Patricia A. Beliefs about academic knowledge. **Educational Psychology Review**, v. 13, n. 4, p. 385-418, 2001.

BUKVOVA, H. Studying Research Collaboration: A Literature Review. Sprouts: Working Papers on Information Systems, 2010. Disponível em: <http://sprouts.aisnet.org/10-3>.

CASPI, Caitlin E. et al. The local food environment and diet: a systematic review. **Health & place**, v. 18, n. 5, p. 1172-1187, 2012.

CASPI, Caitlin E. et al. The local food environment and diet: a systematic review. **Health & place**, v. 18, n. 5, p. 1172-1187, 2012

VANDERWEELE, Tyler J.; VALERI, Linda; OGBURN, CHAVALARIAS, David et al. Evolution of reporting P values in the biomedical literature, 1990-2015. **Jama**, v. 315, n. 11, p. 1141-1148, 2016.

CLEMEN, Gudrun. The concept of hedging: Origins, approaches and definitions. **Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts**, v. 24, p. 235, 1997.

COLE, Stephen R.; FRANGAKIS, Constantine E. The consistency statement in causal inference: a definition or an assumption? **Epidemiology**, v. 20, n. 1, p. 3-5, 2009.

COPE, Mark B.; ALLISON, David B. White hat bias: a threat to the integrity of scientific reporting. **Acta Pædiatrica**, v. 99, n. 11, p. 1615-1617, 2010.

CORTES, Taísa Rodrigues; FAERSTEIN, Eduardo; STRUCHINER, Claudio José. Utilização de diagramas causais em epidemiologia: um exemplo de aplicação em situação de confusão. **Cadernos de Saúde Pública**, v. 32, p. e00103115, 2016.

CROSKERRY, Pat; SINGHAL, Geeta; MAMEDE, Sílvia. Cognitive debiasing 2: impediments to and strategies for change. **BMJ Qual Saf**, 2013a.

CROSKERRY, Pat; SINGHAL, Geeta; MAMEDE, Sílvia. Cognitive debiasing 1: origins of bias and theory of debiasing. **BMJ Qual Saf**, 2013b.

DA COSTA, Bruno R. et al. Uses and misuses of the STROBE statement: bibliographic study. **BMJ open**, v. 1, n. 1, p. e000048, 2011.

DASH, Denver. Caveats for causal reasoning with equilibrium models. 2003. Tese de Doutorado. University of Pittsburgh. Disponível em: <https://pdfs.semanticscholar.org/f462/68c50f815eb31bd86d114f534a71d76f6a54.pdf>.

DAWID, A. Philip. Beware of the DAG! **Journal of Machine Learning Research-Proceedings Track**, v. 6, p. 59-86, 2010.

DEBROCK, Guy; HULSWIT, Menno (Ed.). **Living doubt: essays concerning the epistemology of Charles Sanders Peirce**. Springer Science & Business Media, 2013.

DEEKS, J. J. et al. Evaluating non-randomised intervention studies. **Health technology assessment (Winchester, England)**, v. 7, n. 27, p. iii-x, 1-173, 2003.

DI MARCO, Chrysanne; KROON, Frederick W.; MERCER, Robert E. Using hedges to classify citations in scientific articles. In: **Computing attitude and affect in text: theory and applications**. Springer, Dordrecht, 2006. p. 247-263.

DIDELEZ, Vanessa et al. Graphical models for inference under outcome-dependent sampling. **Statistical Science**, v. 25, n. 3, p. 368-387, 2010.

DOWNEY, H. Kirk; SLOCUM, John W. Uncertainty: Measures, research, and sources of variation. **Academy of Management journal**, v. 18, n. 3, p. 562-578, 1975.

DOWNEY, H. Kirk; HELLRIEGEL, Don; SLOCUM JR, John W. Individual characteristics as sources of perceived uncertainty variability **Human Relations**, v. 30, n. 2, p. 161-174, 1977.

DUMAS, O. et al. Approches d'analyse causale en épidémiologie. **Revue d'Épidémiologie et de Santé Publique**, v. 62, n. 1, p. 53-63, 2014.

DUNCAN, Robert B. Characteristics of organizational environments and perceived environmental uncertainty. **Administrative science quarterly**, p. 313-327, 1972.

DUNNING, David. The Dunning–Kruger effect: On being ignorant of one's own ignorance. In: **Advances in experimental social psychology**. Academic Press, 2011. p. 247-296.

EBRAHIM, Shah; CLARKE, Mike. STROBE: new standards for reporting observational epidemiology, a chance to improve. **International journal of epidemiology**, v. 36, n. 5, p. 946-948, 2007.

ELIZABETH L. The role of measurement error and misclassification in mediation analysis. **Epidemiology (Cambridge, Mass.)**, v. 23, n. 4, p. 561, 2012.

FANELLI, Daniele; COSTAS, Rodrigo; LARIVIÈRE, Vincent. Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. **PLoS One**, v. 10, n. 6, p. e0127556, 2015.

FARKAS, Richárd et al. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In: **Proceedings of the Fourteenth Conference on Computational Natural Language Learning---Shared Task**. Association for Computational Linguistics, 2010. p. 1-12.

FARROKHI, Farahman; EMAMI, Safoora. Hedges and boosters in academic writing: native vs. non-native research articles in applied linguistics and engineering. **Journal of English Language Pedagogy and Practice**, v. 1, n. 2, p. 62-98, 2008.

FESTINGER, Leon. **A theory of cognitive dissonance**. Stanford university press, 1962.

FIELD, Nigel et al. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. **The Lancet Infectious Diseases**, v. 14, n. 4, p. 341-352, 2014.

FIRESTEIN, Stuart. **Ignorance: How It Drives Science**. Oxford University Press, 2012.

FITCHETT, Elizabeth JA et al. Strengthening the Reporting of Observational Studies in Epidemiology for Newborn Infection (STROBE-NI): an extension of the STROBE statement for neonatal infection research. **The Lancet infectious diseases**, v. 16, n. 10, p. e202-e213, 2016.

FOX, Craig R.; ÜLKÜMEN, Gülden; MALLE, Bertram F. On the dual nature of uncertainty: Cues from natural language. **Unpublished manuscript**, 2011.

FRASER, Bruce. Pragmatic competence: The case of hedging. **New approaches to hedging**, v. 1534, 2010.

FLEMING, Stephen M.; DOLAN, Raymond J. The neural basis of metacognitive ability. **Phil. Trans. R. Soc. B**, v. 367, n. 1594, p. 1338-1349, 2012.

FRITH, Chris D. The role of metacognition in human social interactions. **Phil. Trans. R. Soc. B**, v. 367, n. 1599, p. 2213-2223, 2012.

FULLER, Thomas et al. What affects authors' and editors' use of reporting guidelines? Findings from an online survey and qualitative interviews. **PloS one**, v. 10, n. 4, p. e0121585, 2015.

GALEA, Sandro; RIDDLE, Matthew; KAPLAN, George A. Causal thinking and complex system approaches in epidemiology. **International Journal of Epidemiology**, v. 39, n. 1, p. 97-106, 2010.

GALLO, Valentina et al. STrengthening the Reporting of OBservational studies in Epidemiology–Molecular Epidemiology (STROBE-ME): An extension of the STROBE statement. **European journal of clinical investigation**, v. 42, n. 1, p. 1-16, 2012.

GALVAO, Taís Freire; SILVA, Marcus Tolentino; GARCIA, Leila Posenato. Ferramentas para melhorar a qualidade e a transparência dos relatos de pesquisa em saúde: guias de redação científica. **Epidemiol. Serv. Saúde**, Brasília, v. 25, n. 2, p. 427-436, 2016.

GILLAERTS, Paul; VAN DE VELDE, Freek. Interactional metadiscourse in research article abstracts. **Journal of English for Academic purposes**, v. 9, n. 2, p. 128-139, 2010.

GINER-SOROLLA, Roger. Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. **Perspectives on Psychological Science**, v. 7, n. 6, p. 562-571, 2012.

GLYMOUR, M. Maria. Using causal diagrams to understand common problems in social epidemiology. **Methods in social epidemiology**, p. 393-428, 2006.

GLYNN, Adam; QUINN, Kevin. **Non-parametric Mechanisms and Causal Modeling**. Working paper, 2007.

GOUGH, David; OLIVER, Sandy; THOMAS, James (Ed.). **An introduction to systematic reviews**. Sage, 2012.

GRACE, James B. et al. Guidelines for a graph theoretic implementation of structural equation modeling. **Ecosphere**, v. 3, n. 8, p. 1-44, 2012.

GREENLAND, Sander. The effect of misclassification in the presence of covariates. **American journal of epidemiology**, v. 112, n. 4, p. 564-569, 1980.

GREENLAND, Sander. Multiple bias modelling for analysis of observational data. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, v. 168, n. 2, p. 267-306, 2005.

GREENLAND, Sander. Overthrowing the tyranny of null hypotheses hidden in causal diagrams. In: **Heuristics, probability and causality: A tribute to Judea Pearl**, p. 365-382, 2010.

GREENLAND, Sander. For and against methodologies: some perspectives on recent causal and statistical inference debates. **European journal of epidemiology**, v. 32, n. 1, p. 3-20, 2017a.

GREENLAND, Sander. Invited Commentary: The Need for Cognitive Science in Methodology. **American journal of epidemiology**, v. 186, n. 6, p. 639-645, 2017b.

GREENLAND, Sander; GUSTAFSON, Paul. Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction. **American journal of epidemiology**, v. 164, n. 1, p. 63-68, 2006.

GREENLAND, Sander; MANSOURNIA, Mohammad Ali. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. **European journal of epidemiology**, v. 30, n. 10, p. 1101-1110, 2015.

GREENLAND, Sander; PEARL, Judea; ROBINS, James M. Causal diagrams for epidemiologic research. **Epidemiology**, v. 10, n. 1, p. 37-48, 1999

GREENLAND, Sander; PEARL, Judea. Adjustments and their consequences—collapsibility analysis using graphical models. **International Statistical Review**, v. 79, n. 3, p. 401-426, 2011.

GRIES, Stefan Th. Dispersions and adjusted frequencies in corpora. **International journal of corpus linguistics**, v. 13, n. 4, p. 403-437, 2008.

GRIMALDO, Francisco; MARUŠIĆ, Ana; SQUAZZONI, Flaminio. Fragments of peer review: A quantitative analysis of the literature (1969-2015). **PloS one**, v. 13, n. 2, p. e0193148, 2018.

GROENWOLD, R. H. H.; ROVERS, M. M. The Catch-22 of appraisals on the quality of observational studies. **Journal of clinical epidemiology**, v. 63, n. 10, p. 1059-1060, 2010.

GROENWOLD, Rolf HH et al. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. **Annals of epidemiology**, v. 18, n. 10, p. 746-751, 2008.

HAAVELMO, Trygve. The statistical implications of a system of simultaneous equations. **Econometrica, Journal of the Econometric Society**, p. 1-12, 1943.

HACKING, Ian. **The Emergence of Probability: a Philosophical Study of Early Ideas About probability, Induction and Statistical Inference**. 1975.

HALLORAN, M. Elizabeth; STRUCHINER, Claudio J. Study designs for dependent happenings. **Epidemiology**, v. 2, n. 5, p. 331-338, 1991.

HÄMÄLÄINEN, Raimo P. Behavioural issues in environmental modelling—The missing perspective. **Environmental Modelling & Software**, v. 73, p. 244-253, 2015.

HAMRA, Ghassan B.; KAUFMAN, Jay S.; VAHRATIAN, Anjel. Model averaging for improving inference from causal diagrams. **International journal of environmental research and public health**, v. 12, n. 8, p. 9391-9407, 2015.

HARMAN, Gilbert. Knowledge, inference, and explanation. **American Philosophical Quarterly**, v. 5, n. 3, p. 164-173, 1968.

KATZ, J. Sylvan; MARTIN, Ben R. What is research collaboration? **Research policy**, v. 26, n. 1, p. 1-18, 1997.

HERNÁN, Miguel A. et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. **American journal of epidemiology**, v. 155, n. 2, p. 176-184, 2002.

HERNÁN, Miguel A.; COLE, Stephen R. Invited commentary: Causal diagrams and measurement bias. **American Journal of Epidemiology**, v. 170, n. 8, p. 959-962, 2009.

HERNÁN, Miguel A.; HERNANDEZ-DIAZ, Sonia; ROBINS, James M. A structural approach to selection bias. **Epidemiology**, v. 15, n. 5, p. 615-625, 2004.

HERNÁN, Miguel A.; ROBINS, James M. **Causal Inference**. New York: Chapman and Hall/CRC Press, 2016. Disponível em:
<http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/>.

HERNÁN, Miguel A.; ROBINS, James M. Estimating causal effects from epidemiological data. **Journal of epidemiology and community health**, v. 60, n. 7, p. 578-586, 2006.

HERNÁN, Miguel A.; TAUBMAN, Sarah L. Does obesity shorten life? Questioning the importance of well-defined interventions to answer causal questions. **International Journal of Obesity**, v. 32, p. S8-S14, 2008.

HIGGINSON, Andrew D.; MUNAFÒ, Marcus R. Current incentives for scientists lead to underpowered studies with erroneous conclusions. **PLoS biology**, v. 14, n. 11, 2016.

HINKEL, Eli (Ed.). **Culture in second language teaching and learning**. Cambridge University Press, 1999.

HODSON, Gordon; SORRENTINO, Richard M. Uncertainty orientation and the Big Five personality structure. **Journal of Research in Personality**, v. 33, n. 2, p. 253-261, 1999.

HOLLAND, Paul W. Statistics and causal inference. **Journal of the American statistical Association**, v. 81, n. 396, p. 945-960, 1986.

HUANG, Mu Hsuan; LIN, Chi-Shiou; CHEN, Dar- Zen. Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact. **Journal of the Association for Information Science and Technology**, v. 62, n. 12, p. 2427-2436, 2011.

KUROKI, Manabu; PEARL, Judea. Measurement bias and effect restoration in causal inference. **Biometrika**, v. 101, n. 2, p. 423-437, 2014.

HYLAND, Ken. **Hedging in scientific research articles**. John Benjamins Publishing, 1998.

HYLAND, Ken; TSE, Polly. Metadiscourse in academic writing: A reappraisal. **Applied linguistics**, v. 25, n. 2, p. 156-177, 2004.

IOANNIDIS, John PA. Limitations are not properly acknowledged in the scientific literature. **Journal of clinical epidemiology**, v. 60, n. 4, p. 324-329, 2007.

IOANNIDIS, John PA. Meta-research: Why research on research matters. **PLoS biology**, v. 16, n. 3, p. e2005468, 2018.

IOANNIDIS, John PA et al. Meta-research: evaluation and improvement of research methods and practices. **PLoS biology**, v. 13, n. 10, p. e1002264, 2015.

JANIS, Irving L. Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. **Boston: Houghton Mifflin**, v. 3, p. 4, 1972.

JEFFERSON, Tom et al. Effects of editorial peer review: a systematic review. **Jama**, v. 287, n. 21, p. 2784-2786, 2002.

JIANG, Zhichao; DING, Peng. The directions of selection bias. **Statistics & Probability Letters**, v. 125, p. 104-109, 2017.

JONNALAGADDA, Siddhartha R.; GOYAL, Pawan; HUFFMAN, Mark D. Automating data extraction in systematic reviews: a systematic review. **Systematic reviews**, v. 4, n. 1, p. 78, 2015.

JUREK, Anne M. et al. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. **European journal of epidemiology**, v. 21, n. 12, p. 871-876, 2006.

JUREK, Anne M. et al. Proper interpretation of non-differential misclassification effects: expectations vs observations. **International journal of epidemiology**, v. 34, n. 3, p. 680-687, 2005.

KAHNEMAN, Daniel, & TVERSKY, Amos. The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) **Judgement under uncertainty: Heuristic and biases**. New York: Cambridge University Press, pp.201-208, 1982b.

KAHNEMAN, Daniel, & TVERSKY, Amos. Variants of uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) **Judgement under uncertainty: Heuristic and biases**. New York: Cambridge University Press, pp.201-208, 1982a

KARAHALIOS, Amalia et al. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. **BMC medical research methodology**, v. 12, n. 1, p. 1, 2012.

KARVANEN, Juha. Study design in causal models. **Scandinavian Journal of Statistics**, v. 42, n. 2, p. 361-377, 2015.

KHANGURA, Sara et al. Evidence summaries: the evolution of a rapid review approach. **Systematic reviews**, v. 1, n. 1, p. 1, 2012.

KIIVERI, H.; SPEED, T. **Structural analysis of multivariate data: A review**. Sociological Methodology, Leinhardt, S. (Ed.). Jossey-Bass, San Francisco, 1982.

KILICOGLU, Halil; BERGLER, Sabine. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. **BMC bioinformatics**, v. 9, n. 11, p. S10, 2008.

KIMMEL, Michael S. **The gendered society**. New York: Oxford University. Press, 2000

KINCAID, Harold. Causal modelling, mechanism, and probability in epidemiology. **Causality in the Sciences**, p. 70-90, 2011.

KIRITCHENKO, Svetlana et al. ExaCT: automatic extraction of clinical trial characteristics from journal publications. **BMC medical informatics and decision making**, v. 10, n. 1, p. 56, 2010.

KITAEV-SMYK, L. A. Factors tensions of the creative process. **Questions of psychology**, n. 3, p. 69-82, 2007.

KNÜPPEL, Sven; STANG, Andreas. DAG Program: Identifying Minimal Sufficient Adjustment Sets. **Epidemiology**, v. 21, n. 1, p. 159, 2010.

KORNILOVA, Tatiana V.; KORNILOV, Sergey A. Intelligence and tolerance/intolerance for uncertainty as predictors of creativity. **Psychology in Russia: State of the Art**, v. 3, p. 240-256, 2010.

KOURILOVA, M. Some linguistic and cultural aspects of scientific discourse processing. **Vienna English Working Papers**, v. 4, n. 2, p. 120-130, 1995.

KOVANIS, Michail et al. Evaluating alternative systems of peer review: a large-scale agent-based modelling approach to scientific publication. **Scientometrics**, v. 113, n. 1, p. 651-671, 2017.

HOWARD, Brian E. et al. SWIFT-Review: a text-mining workbench for systematic review. **Systematic reviews**, v. 5, n. 1, p. 87, 2016.

KRIEGER, Nancy; SMITH, George Davey. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. **International Journal of Epidemiology**, p. dyw114, 2016.

KRUEGER, Joachim; MUELLER, Ross A. Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. **Journal of personality and social psychology**, v. 82, n. 2, p. 180, 2002.

KRUGER, Justin; DUNNING, David. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. **Journal of personality and social psychology**, v. 77, n. 6, p. 1121, 1999.

KUNDA, Ziva. The case for motivated reasoning. **Psychological bulletin**, v. 108, n. 3, p. 480, 1990.

LACHAT, Carl et al. Strengthening the Reporting of Observational Studies in Epidemiology—Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement. **PLoS Med**, v. 13, n. 6, p. e1002036, 2016.

LAKOFF, George. **Hedges: A study in meaning criteria and the logic of fuzzy concepts**. Chicago Linguistic Society 8: 182-222, 1987.

LAKOFF, Robin. Language and woman's place. **Language in society**, v. 2, n. 1, p. 45-79, 1973.

LANG, Shona; KLEIJNEN, Jos. Quality assessment tools for observational studies: lack of consensus. **International journal of evidence-based healthcare**, v. 8, n. 4, p. 247, 2010.

LASH, Timothy L. Bias analysis applied to Agricultural Health Study publications to estimate non-random sources of uncertainty. **Journal of Occupational Medicine and Toxicology**, v. 2, n. 1, p. 1, 2007.

LASH, Timothy L.; FOX, Matthew P.; FINK, Aliza K. **Applying quantitative bias analysis to epidemiologic data**. Springer Science & Business Media, 2011.

LEAPER, Campbell; ROBNETT, Rachael D. Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. **Psychology of Women Quarterly**, v. 35, n. 1, p. 129-142, 2011.

LITTLE, Julian et al. STrengthening the REporting of Genetic Association studies (STREGA)—an extension of the STROBE statement. **European journal of clinical investigation**, v. 39, n. 4, p. 247-266, 2009.

LU, Zhiyong. PubMed and beyond: a survey of web tools for searching biomedical literature. **Database**, v. 2011, p. baq036, 2011.

LUIZ, Ronir Raggio; STRUCHINER, Claudio José. **Inferência Causal em Epidemiologia: o modelo de respostas potenciais**. Editora Fiocruz, 2002.

MALTA, Monica et al. STROBE initiative: guidelines on reporting observational studies. **Revista de Saúde Pública**, v. 44, n. 3, p. 559-565, 2010.

MANSOURNIA, Mohammad A.; HERNÁN, Miguel A.; GREENLAND, Sander. Matched designs and causal diagrams. **International journal of epidemiology**, v. 42, n. 3, p. 860-869, 2013.

MARUŠIĆ, Ana; MARUŠIĆ, Matko. Teaching students how to read and write science: a mandatory course on scientific research and communication in medicine. **Academic medicine**, v. 78, n. 12, p. 1235-1239, 2003.

MIRZAPOUR, Fatemeh; MAHAND, Mohammad Rasekh. Hedges and boosters in native and non-native library and information and computer science research articles. **3L: Language, Linguistics, Literature**, v. 18, n. 2, 2012.

MONTIBELLER, Gilberto; WINTERFELDT, Detlof. Cognitive and motivational biases in decision and risk analysis. **Risk Analysis**, v. 35, n. 7, p. 1230-1251, 2015.

MONTIBELLER, Gilberto; VON WINTERFELDT, Detlof. Individual and group biases in value and uncertainty judgments. In: **Elicitation**. Springer, Cham, p. 377-392, 2018.

MORLAND, Kimberly B. (Ed.). **Local food environments: food access in America**. Crc Press, 2014.

MURAD, Mohammad Hassan; WANG, Zhen. Guidelines for reporting meta-epidemiological methodology research. **BMJ evidence-based medicine**, 2017.

NEYMAN, Jerzy. On the application of probability theory to agricultural experiment: essay on principles, 1923. Reimpresso em inglês em **Statistical Science**, v. 5, n. 4, p. 465-472, 1990.

NICKERSON, Raymond S. Confirmation bias: A ubiquitous phenomenon in many guises. **Review of general psychology**, v. 2, n. 2, p. 175, 1998.

NGUYEN, Trang Quynh; DAFOE, Allan; OGBURN, Elizabeth L. Collider bias in binary variable structures. **arXiv preprint arXiv:1609.00606**, 2016.

NOLA, Robert; SANKEY, Howard. **Theories of scientific method: an introduction**. Routledge, 2014.

NOSEK, Brian A.; SPIES, Jeffrey R.; MOTYL, Matt. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. **Perspectives on Psychological Science**, v. 7, n. 6, p. 615-631, 2012.

NUYTS, Jan. **Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective**. John Benjamins Publishing, 2001.

O'CONNOR, A. M. et al. Explanation and Elaboration Document for the STROBE Vet Statement: Strengthening the Reporting of Observational Studies in Epidemiology—Veterinary Extension. **Journal of Veterinary Internal Medicine**, 2016.

OGBURN, Elizabeth L. et al. Causal diagrams for interference. **Statistical science**, v. 29, n. 4, p. 559-578, 2014.

OGBURN, Elizabeth L.; VANDERWEELE, Tyler J. On the nondifferential misclassification of a binary confounder. **Epidemiology (Cambridge, Mass.)**, v. 23, n. 3, p. 433, 2012.

PAPATHANASIOU, Afroditi A.; ZINTZARAS, Elias. Assessing the quality of reporting of observational studies in cancer. **Annals of epidemiology**, v. 20, n. 1, p. 67-73, 2010.

PARK, Won Woo. A review of research on groupthink. **Journal of Behavioral Decision Making**, v. 3, n. 4, p. 229-245, 1990.

PASTORINO, Roberta et al. Quality assessment of studies published in open access and subscription journals: results of a systematic evaluation. **PloS One**, v. 11, n. 5, p. e0154217, 2016.

PEARL, Judea. Causal diagrams for empirical research. **Biometrika**, v. 82, n. 4, p. 669-688, 1995.

PEARL, Judea. **Causality: models, reasoning and inference**. Cambridge: MIT press, 2000.

PEARL, Judea. An introduction to causal inference. **The international journal of biostatistics**, v. 6, n. 2, 2010a.

PEARL, Judea. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? *Epidemiology*, v. 21, n. 6, p. 872-875, 2010b.

PEARL, Judea. The causal mediation formula—a guide to the assessment of pathways and mechanisms. **Prevention Science**, v. 13, n. 4, p. 426-436, 2012.

PEARL, Judea. Comment: understanding Simpson's paradox. **The American Statistician**, v. 68, n. 1, p. 8-13, 2014a.

PEARL, Judea. Is scientific knowledge useful for policy analysis? a peculiar theorem says: No. **Journal of Causal Inference**. v. 2, n. 1, p. 109-112, 2014b.

PEARL, Judea; BAREINBOIM, Elias. External validity: From do-calculus to transportability across populations. **Statistical Science**, v. 29, n. 4, p. 579-595, 2014.

PEARL, Judea; GLYMOUR, Madelyn; JEWELL, Nicholas P. Causal inference in statistics: a primer. **John Wiley & Sons**, 2016.

PEARL, Judea; ROBINS, James. Probabilistic evaluation of sequential plans from causal models with hidden variables. In: **Proceedings of the Eleventh conference on Uncertainty in artificial intelligence**. Morgan Kaufmann Publishers Inc., 1995. p. 444-453.

PEARSON, Karl. **The grammar of science**. 1892.

PEIRCE, Charles Sanders; HARTSHORNE, Charles; WEISS, Paul. **Collected Papers of Charles Sanders Peirce**: Vol. III, Exact Logic. 1935.

PENFIELD, Teresa et al. Assessment, evaluations, and definitions of research impact: A review. **Research Evaluation**, v. 23, n. 1, p. 21-32, 2014.

PEREIRA, Tiago V.; HORWITZ, Ralph I.; IOANNIDIS, John PA. Empirical evaluation of very large treatment effects of medical interventions. **Jama**, v. 308, n. 16, p. 1676-1684, 2012.

PETERSEN, Maya L.; VAN DER LAAN, Mark J. Causal models and learning from data: integrating causal modeling and statistical estimation. **Epidemiology (Cambridge, Mass.)**, v. 25, n. 3, p. 418, 2014.

PINKER, Steven. Why academics stink at writing. **The chronicle of higher education**, v. 26, 2014.

POCOCK, Stuart J. et al. Issues in the reporting of epidemiological studies: a survey of recent practice. **BmJ**, v. 329, n. 7471, p. 883, 2004.

PORTA, Miquel; VINEIS, Paolo; BOLÚMAR, Francisco. The current deconstruction of paradoxes: one sign of the ongoing methodological "revolution". **European journal of epidemiology**, v. 30, n. 10, p. 1079-1087, 2015.

POUWELS, Koen B. et al. Quality of reporting of confounding remained suboptimal after the STROBE guideline. **Journal of clinical epidemiology**, v. 69, p. 217-224, 2016.

PROKOFIEVA, Anna; HIRSCHBERG, Julia. Hedging and speaker commitment. **Emotion, social signals, sentiment & linked open data**, p. 10, 2014.

PUHAN, M. A. et al. Acknowledging limitations in biomedical studies: The ALIBI study. In: **The Sixth International Congress on Peer Review and Biomedical Publication**. 2009. p. 10-12.

PUHAN, Milo A. et al. Discussing study limitations in reports of biomedical studies- the need for more transparency. **Health and quality of life outcomes**, v. 10, n. 1, p. 1, 2012.

RADHAKRISHNAN, Adityanarayanan; SOLUS, Liam; UHLER, Caroline. Counting Markov equivalence classes by number of immoralities. In: **33rd Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, 11 August 2017 through 15 August 2017**. AUAI Press Corvallis, 2017.

RAMAKRISHNAN, Cartic et al. Layout-aware text extraction from full-text PDF of scientific articles. **Source code for biology and medicine**, v. 7, n. 1, p. 7, 2012.

REGAN, Helen M.; COLYVAN, Mark; BURGMAN, Mark A. A taxonomy and treatment of uncertainty for ecology and conservation biology. **Ecological applications**, v. 12, n. 2, p. 618-628, 2002.

REICHENHEIM, Michael E.; MORAES, Claudia Leite. Alguns pilares para a apreciação da validade de estudos epidemiológicos. **Rev Bras Epidemiol**, v. 1, n. 2, p. 131-48, 1998.

ROBINS, James M. Data, design, and background knowledge in etiologic inference. **Epidemiology**, v. 12, n. 3, p. 313-320, 2001.

ROBINS, James. 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**, v. 7, n. 9-12, p. 1393-1512, 1986.

ROLAND, Marie Claude. Publish and perish. **EMBO reports**, v. 8, n. 5, p. 424-428, 2007.

ROSE, James D. Diverse perspectives on the groupthink theory—a literary review. **Emerging Leadership Journeys**, v. 4, n. 1, p. 37-57, 2011.

ROTHMAN, Kenneth J. **Causal inference**. Epidemiology Resources Inc., 1988

ROTHMAN, Kenneth J.; GREENLAND, Sander. Causation and causal inference in epidemiology. **American journal of public health**, v. 95, n. S1, p. S144-S150, 2005.

ROTHMAN, Kenneth J.; POOLE, Charles. Some guidelines on guidelines: they should come with expiration dates. **Epidemiology**, v. 18, n. 6, p. 794-796, 2007.

RUBIN, Donald B. Comment: Neyman (1923) and causal inference in experiments and observational studies. **Statistical Science**, v. 5, n. 4, p. 472-480, 1990.

RUBIN, Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of educational Psychology**, v. 66, n. 5, p. 688, 1974.

RUBIN, Donald B. Inference and missing data. **Biometrika**, v. 63, n. 3, p. 581-592, 1976.

SALAGER-MEYER, Françoise. Hedges and textual communicative function in medical English written discourse. **English for specific purposes**, v. 13, n. 2, p. 149-170, 1994.

SAMAAN, Zainab et al. A systematic scoping review of adherence to reporting guidelines in health care literature. **Journal of multidisciplinary healthcare**, v. 6, p. 169, 2013.

SANDERSON, Simon; TATT, Iain D.; HIGGINS, Julian PT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. **International journal of epidemiology**, v. 36, n. 3, p. 666-676, 2007.

SÁNCHEZ-NIUBÒ, Albert; G FORERO, Carlos; G BENAVIDES, Fernando. Utilidad de los diagramas causales para conceptualizar mecanismos en epidemiología laboral. **Archivos de Prevención de Riesgos Laborales**, v. 19, n. 2, p. 103-106, 2016.

SCHIPF, S. et al. Directed Acyclic Graphs (DAGs)—Die Anwendung kausaler Graphen in der Epidemiologie. **Das Gesundheitswesen**, v. 73, n. 12, p. 888-892, 2011.

SCHNEIDER, Stephen H.; KUNTZ-DURISETI, Kristin. Uncertainty and climate change policy. **Climate change policy: a survey**, p. 53-87, 2002.

SENOS, Jorge. **Pensamento contrafactual e raciocínio causal: Efeito de facilitação recíproca e modelo de integração**. 2008 Tese de doutorado. Universidade Nova de Lisboa: Instituto Superior de Psicologia Aplicada. Disponível em: <http://hdl.handle.net/10400.12/207>

SHAHAR, Eyal; SHAHAR, Doron J. Causal diagrams, information bias, and thought bias. **Pragmatic and Observational Research**, v. 1, p. 33, 2010.

SHAHAR, Eyal; SHAHAR, Doron J. Causal diagrams, information bias, and thought bias. **Pragmatic and Observational Research**, v. 1, p. 33, 2010.

SHARP, Stephen J. et al. A review of published analyses of case-cohort studies and recommendations for future reporting. **PloS one**, v. 9, n. 6, p. e101176, 2014.

SHPITSER, Ilya; PEARL, Judea. Complete identification methods for the causal hierarchy. **The Journal of Machine Learning Research**, v. 9, p. 1941-1979, 2008.

SHPITSER, Ilya. **Complete Identification Methods for Causal Inference**. Tese de Doutorado. UCLA Cognitive Systems Laboratory, 2008. Disponível em: ftp.cs.ucla.edu/pub/stat_ser/shpitser-thesis.pdf.

SHRIER, Ian; PLATT, Robert W. Reducing bias through directed acyclic graphs. **BMC medical research methodology**, v. 8, n. 1, p. 70, 2008.

SJÖLANDER, Arvid; GREENLAND, Sander. Ignoring the matching variables in cohort studies—when is it valid and why?. **Statistics in medicine**, v. 32, n. 27, p. 4696-4708, 2013.

SKELTON, John. The care and maintenance of hedges. **ELT journal**, v. 42, n. 1, p. 37-43, 1988.

SLOMAN, Steven. **Causal models: How people think about the world and its alternatives**. Oxford University Press, 2005.

SMALDINO, Paul E.; MCELREATH, Richard. The natural selection of bad science. **Royal Society Open Science**, v. 3, n. 9, p. 160384, 2016.

SMITHSON, Michael. **Ignorance and uncertainty: emerging paradigms**. Springer Science & Business Media, 1989.

SPIRITES, Peter; GLYMOUR, Clark; SCHEINES, Richard. **Causation, prediction, and search**. The MIT Press, 2000.

STERNE, Jonathan AC et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. **Bmj**, v. 338, p. b2393, 2009.

STEVENS, Adrienne et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. **Bmj**, v. 348, p. g3804, 2014.

SUZUKI, Etsuji et al. [Causal Inference in Medicine Part II. Directed acyclic graphs--a useful method for confounder selection, categorization of potential biases, and hypothesis specification]. **Nihon eiseigaku zasshi. Japanese journal of hygiene**, v. 64, n. 4, p. 796-805, 2009.

SZKLO, Moyses. Epidemiologia translacional: algumas considerações. **Epidemiologia e Serviços de Saúde**, 24(1): 161-172, 2015.

TATIONI, Athina; BONITSIS, Nikolaos G.; IOANNIDIS, John PA. Persistence of contradicted claims in the literature. **Jama**, v. 298, n. 21, p. 2517-2526, 2007.

TER RIET, Gerben et al. All that glitters isn't gold: A survey on acknowledgment of limitations in biomedical studies. **PloS one**, v. 8, n. 11, p. e73623, 2013.

TEXTOR, Johannes; HARDT, Juliane; KNÜPPEL, Sven. DAGitty: a graphical tool for analyzing causal diagrams. **Epidemiology**, v. 22, n. 5, p. 745, 2011.

THOMAS, James et al. Living systematic reviews: 2. Combining human and machine effort. **Journal of clinical epidemiology**, v. 91, p. 31-37, 2017.

THOMAS, Rick P. et al. Diagnostic hypothesis generation and human judgment. **Psychological review**, v. 115, n. 1, p. 155, 2008.

THOMPSON, Caroline A.; ARAH, Onyebuchi A. Selection bias modeling using observed data augmented with imputed record-level probabilities. **Annals of epidemiology**, v. 24, n. 10, p. 747-753, 2014.

TIAN, Jin; PEARL, Judea. On the testable implications of causal models with hidden variables. In: **Proceedings of the 18th conference on uncertainty in artificial intelligence**. Morgan Kaufmann Publishers Inc., p. 519-527, 2002.

TIAN, Jin. **Studies in Causal Reasoning and Learning**. 2002. Tese de Doutorado. University of California Los Angeles, CA 90095-1596, USA.

TORVIK, Vetle I.; AGARWAL, Sneha. Ethnea -- an instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database. **International Symposium on Science of Science March**. Library of Congress, 2016.

TRUEGER, N. Seth et al. The altmetric score: a new measure for article-level dissemination and impact. **Annals of emergency medicine**, v. 66, n. 5, p. 549-553, 2015.

TURNER, Lucy et al. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review a. **Systematic reviews**, v. 1, n. 1, p. 1, 2012.

TURNER, Lucy et al. Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. **Systematic reviews**, v. 1, n. 1, p. 60, 2012.

TVERSKY, Amos; KAHNEMAN, Daniel. Advances in prospect theory: Cumulative representation of uncertainty. **Journal of Risk and uncertainty**, v. 5, n. 4, p. 297-323, 1992.

TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty: Heuristics and biases. **Science**, v. 185, n. 4157, p. 1124-1131, 1974.

TSAFNAT, Guy et al. Systematic review automation technologies. **Systematic reviews**, v. 3, n. 1, p. 74, 2014.

ÜLKÜMEN, Gülden; FOX, Craig R.; MALLE, Bertram F. Two dimensions of subjective uncertainty: Clues from natural language. **Journal of Experimental Psychology: General**, v. 145, n. 10, p. 1280, 2016.

VAN ECK, Nees Jan et al. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. **Journal of the Association for Information Science and Technology**, v. 61, n. 12, p. 2405-2416, 2010.

VAN HOECK, Nicole; WATSON, Patrick D.; BARBEY, Aron K. Cognitive neuroscience of human counterfactual reasoning. **Frontiers in human neuroscience**, v. 9, 2015.

VANDENBROUCKE, Jan P. et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. **Gaceta Sanitaria**, v. 23, n. 2, p. 158e1-158e28, 2009.

VANDENBROUCKE, Jan P.; BROADBENT, Alex; PEARCE, Neil. Causality and causal inference in epidemiology: the need for a pluralistic approach. **International journal of epidemiology**, p. dyv341, 2016.

VANDERWEELE, Tyler J. Concerning the consistency assumption in causal inference. **Epidemiology**, v. 20, n. 6, p. 880-883, 2009.

VANDERWEELE, Tyler J.; DING, Peng. Sensitivity analysis in observational research: introducing the E-value. **Annals of internal medicine**, v. 167, n. 4, p. 268-274, 2017.

VARTTALA, Teppo. **Hedging in scientifically oriented discourse. Exploring variation according to discipline and intended audience**. 2001. Tese de doutorado. Tampere University. 2001. Disponível em: <http://acta.uta.fi/pdf/951-44-5195-3.pdf>.

VELDKAMP, Coosje LS et al. Who believes in the storybook image of the scientist? **Accountability in research**, v. 24, n. 3, p. 127-151, 2017.

VERMA, T.; PEARL, Judea. Equivalence and synthesis of causal models. In: **Proceedings of the Sixth annual conference on uncertainty in artificial Intelligence**, Vol. 6, p. 220-227, Elsevier Science Publishers, Cambridge, MA, 1991.

VON ELM, Erik et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. **Preventive medicine**, v. 45, n. 4, p. 247-251, 2007.

WESTREICH, Daniel; GREENLAND, Sander. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. **American journal of epidemiology**, v. 177, n. 4, p. 292-298, 2013.

WILLIAMSON, Jon. Causal Pluralism versus Epistemic Causality. **Philosophica**, v. 77, p. 69-96, 2006.

WITTGENSTEIN, Ludwig et al. **On certainty**. Oxford: Blackwell, 1969.

ZHENG, YJ; ZHAO, N. [Directed acyclic graphs: languages, rules and applications]. **Zhonghua Liu Xing Bing Xue Za Zhi**, 10;38 (8):1140-1144, 2017

APÊNDICE A - Estratégias de buscas

Base			
Diretório PubMed Central			
Atualização 29/01/2018—Horário: 22:00			
	Termos	Outros termos	Itens
1	Directed acyclic graph*[ARTICLE] OR directed acyclic graph*[CAPT]	Causal directed acyclic graph	3915
2	Causal network*[ARTICLE] OR Causal network*[CAPT]	Causal network diagram, causal network model	778
3	Causal diagram*[ARTICLE] OR causal diagram* [CAPT]		584
4	DAG[CAPT] AND causal [CAPT]		120
5	Causal graph*[ARTICLE] OR causal graph*[CAPT]	Directed acyclic causal graph, causal graph model, directed causal graph	245
6	Causal bayesian network*[ARTICLE] OR Causal bayesian network*[CAPT]		55
7	Structural causal model* [ARTICLE] OR structural causal model* [CAPT]		28
8	Causal path diagram* [All Fields]		3
9	Causal model diagram* [All Fields]		1
10	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7		5168
11	"2000/01/01"[PDAT]: "2017/12/31"[PDAT]		
12	10 AND 11		4988
13	"Gene ontology"[MeSH Terms] OR Gene ontology[Text Word] OR "Algorithms"[Title] OR "Natural Language Processing"[Mesh]		
14	12 NOT 13		3770

Base		Medline (via PubMed)	
		Atualização 29/01/2018—Horário: 20:25	
1	Directed acyclic graph*[All Fields]		495
2	Causal network*[All Fields]		212
3	Causal diagram*[All Fields]		121
4	Causal graph*[All Fields]		59
5	Structural causal model*[All Fields]		11
6	(causal[All Fields] AND path diagram*[All Fields])		15
7	1 OR 2 OR 3 OR 4 OR 5 OR 6		897
8	"2000/01/01"[PDAT]: "2017/12/31"[PDAT]		
9	7 AND 8		845
10	"Gene ontology"[MeSH Terms] OR Gene ontology[Text Word] OR "Algorithms"[Title] OR "Natural Language Processing"[Mesh]		
11	9 NOT 10		796
Base		Web of Science	
		Atualização 29/01/2018—Horário: 22:51	
Etapa	Termos	Outros termos englobados	Itens
1	TS=(directed NEAR/1 acyclic NEAR/1 graph*)	directed acyclic causal graph, causal directed acyclic graph	3465
2	TS=(causal NEAR/1 diagram*)	Causal model diagram, causal path diagram, causal network diagram,	787
3	TS=(causal NEAR/1 network*)	Causal bayesian network, causal network model	901
4	TS=(structural NEAR/1 causal NEAR/1 model*)		52

5	TS=(graphical NEAR/1 causal NEAR/1 model*)		44
6	TS=(causal NEAR/0 DAG)		1
7	1 OR 2 OR 3 OR 4 OR 5 OR 6		851
8	PY=2000-2017		
9	Document Types: Review OR Book OR Case Report OR Meeting OR Editorial OR Letter OR News OR Patent OR Clinical Trial OR Report OR Correction		
10	7 AND 8		5949
11	10 NOT 9		2833
12	TS=(Gene ontology OR Natural Language Processing) OR TI= Algorithms		
13	11 NOT 12		2601

APÊNDICE B - Questionário

N	Item	Descrição	Opções
	PMID	1º Número PDF	
	Número	2º Número PDF	
<u>Características gerais do estudo</u>			
1	Desenho do estudo		1) Estudo transversal 2) Estudo longitudinal 3) Estudo caso controle 4) Outro (especificar)
2	Objetivo do estudo		1) Análise do efeito total 2) Análise de mediação 3) Modificação de efeito 4) Outro (especificar)
3	Tipo de exposição de interesse		1) Fixa 2) Tempo-depentende 3) Múltiplas variáveis
	Variável de exposição		
4	Tipo de desfecho		1) Fixo 2) Tempo-depentende 3) Múltiplas variáveis
	Variável de desfecho		
5	Outra variável de interesse (tipo)		1) Mediador 2) Modificador de

			efeito 3) Variável instrumental 4) Outros
6	País do estudo		
7	Tamanho da amostra		
<u>Relato de confundimento</u>			
	Os autores mencionaram as razões pelas quais os potenciais confundidores foram selecionados para as análises?		1) Sim 2) Não 3) Incerto 4) Não se aplica
2	Qual foi a estratégia de seleção dos potenciais confundidores?		1) Revisão da literatura/critérios teóricos 2) DAG 3) Outro (especificar)
3	Os autores mencionaram as razões para inclusão dos confundidores no modelo final?		1) Sim 2) Não 3) Incerto 4) Não se aplica
4	Quais foram os critérios	Modelo total ou DAG <i>modelo total: todos os confundidores selecionados previamente foram incluídos no modelo final. Ou seja,</i>	1) Modelo total 2) DAG modelo total 3) Significância estatística (CS) 4) Mudança na

	de seleção dos confundidores para o modelo final?	<i>apenas critérios teóricos de seleção e nenhum outro critério de inclusão ou exclusão de covariáveis foi descrito.</i>	estimativa (CIE) 5) CIE e CS 6) Seleção stepwise. 7) Outro (especificar)
5	Quantos confundidores foram incluídos no modelo final?		
6	Os autores apresentaram as estimativas ajustadas para as exposições e confundidores a partir de um único modelo?		1) Sim 2) Não 3) Incerto 4) Não se aplica
7	Qual foi a interpretação das estimativas de efeitos dos confundidores?	Sem interpretação: os resultados não foram discutidos Mesma interpretação: todas as estimativas foram descritas da mesma forma (p.ex. associação, relação, efeito) Relataram as diferenças entre as estimativas: os autores fizeram algum tipo de distinção entre as estimativas (efeito total, direto, mediação)	1) Sem interpretação 2) Mesma interpretação 3) Relataram as diferenças 4) Outros comentários 5) Não se aplica
8	Os autores relataram o método utilizado para o		1) Sim 2) Não

	controle de confundimento?		3) Incerto 4) Não se aplica
9	Qual foi o método utilizado para o controle de confundimento?		
10	Os autores relataram alguma análise de sensibilidade para avaliar o confundimento não observado?	<i>Considere qualquer tipo de análise descrita pelos autores com o objetivo de avaliar o impacto da limitação ou viés.</i>	1) Sim 2) Não 3) Incerto 4) Não se aplica
11	Qual o tipo de análise de sensibilidade realizada?	<p>Análise de sensibilidade (tradicional): <i>em geral, estratégias envolvendo a inclusão ou exclusão de variáveis no modelo, restrição da amostra ou redefinição de variáveis.</i></p> <p>Formula de vies (Bias-adjustment formulas): <i>calcular algebricamente a magnitude do viés necessária para explicar a associação entre E e D.</i></p> <p>Modelagem de viés: <i>Probabilistic bias analysis, Bayesian sensitivity analysis, Monte Carlo sensitivity analysis: técnicas semelhantes que utilizam</i></p>	1) Análise de sensibilidade tradicional 2) Formula de vies 3) Modelagem de viés, (especificar)

		<i>uma distribuição dos parâmetros do viés – estipulados a priori - para reajustar as medidas de efeito e precisão.</i>	
12	Os autores apresentaram as características dos principais fatores de confusão?		1) Sim 2) Não 3) Incerto 4) Não se aplica
13	Os autores apresentaram as estimativas brutas e as estimativas ajustadas?	<i>Marque como “incerto”, se há dados para o cálculo das medidas brutas (p.ex. frequência das categorias do desfecho entre expostos e não expostos)</i>	1) Sim 2) Não 3) Incerto 4) Não se aplica
<u>Relato das limitações do estudo</u>			
1	Os autores reconheceram alguma limitação do estudo ou a possibilidade de viés?	<i>Considere todas as limitações ou vieses mencionados pelos autores, em qualquer parte do artigo (não apenas na discussão ou na seção das limitações).</i>	1) Sim 2) Não 3) Incerto 4) Não se aplica
2	Qual foi o tipo de	<i>Classificar como “erros de mensuração” qualquer comentário relativo a mensuração, classificação ou definição de variáveis.</i>	1) Confundimento não observado ou residual 2) Confundimento entre M-D 4) Outras variáveis

	limitação ou viés em potencial relatado pelos autores?		<p>não observadas</p> <p>5) Erros de mensuração da exposição</p> <p>6) Erros de mensuração do desfecho</p> <p>7) Erros de mensuração do mediador</p> <p>8) Erros de mensuração (geral)</p> <p>9) Dados faltantes</p> <p>10) Perda de seguimento</p> <p>11) Viés de seleção (geral)</p> <p>12) Tempo</p> <p>13) Validade externa</p> <p>14) Tamanho da amostra (poder, precisão)</p> <p>15) Outra (especificar)</p>
		<i>Descreva o tipo de avaliação da limitação para cada limitação ou viés relatado</i>	
3	Os autores discutiram qualitativamente a possível direção do viés?		<p>1) Sim</p> <p>2) Não</p> <p>3) Incerto</p> <p>4) Não se aplica</p>

4	Os autores avaliaram quantitativamente o possível impacto do viés/limitação?	<p>Análise de sensibilidade (tradicional): <i>em geral, estratégias envolvendo a inclusão ou exclusão de variáveis no modelo, restrição da amostra ou redefinição de variáveis.</i></p> <p>Formula de vies (Bias-adjustment formulas): <i>calcular algebricamente a magnitude do viés necessária para explicar a associação entre E e D.</i></p> <p>Modelagem de viés: <i>Missing data methods – técnicas de imputação de dados, IPW- Inverse probability weighting (IPCW, IPTW), Probabilistic bias analysis, Bayesian sensitivity analysis, Monte Carlo sensitivity analysis: técnicas semelhantes que utilizam uma distribuição dos parâmetros do viés – estipulados a priori - para reajustar as medidas de efeito e precisão.</i></p>	<p>1) Sim, análise de sensibilidade (especificar)</p> <p>2) Sim, modelagem de viés, (especificar)</p> <p>3) Não</p>
---	--	--	---

1	O DAG foi apresentado no artigo?		1) Sim, no artigo 2) Sim, no material suplementar 3) Não
2	Os autores apresentaram o referencial teórico para o DAG?		1) Sim, comentários com referências 2) Parcialmente, comentários gerais 3) Não
3	Os autores discutiram alguma incerteza sobre a estrutura gráfica?	<i>Considere apenas os comentários direcionados (explícitos) à estrutura gráfica. Desconsidere as discussões gerais sobre a possibilidade de confundimento, erros de mensuração etc.</i>	1) Sim 2) Não 3) Incerto
4	Qual foi o tipo de incerteza relatada pelos autores?	<i>Se necessário, marque mais de uma opção</i>	1) Incerteza sobre a presença de vértices 2) Incerteza sobre a presença de setas 3) Incerteza sobre a direção das setas 4) Ciclos causais ou relações bidirecionais 5) Incerteza sobre a função de um vértice 6) Outras (especificar)
5	Os autores discutiram como essa incerteza poderia afetar os		1) Sim 2) Não 3) Incerto

	resultados?		
6	Quantos DAG foram apresentados?		
7	Um único DAG foi utilizado para guiar as análises de diferentes exposições e/ou desfechos?		1) Sim 2) Não 3) Incerto
8	Qual foi o motivo?		
9	Algum DAG apresenta variáveis agrupadas em um único vértice?		1) Sim 2) Não 3) Incerto
10	Quantos vértices possuem variáveis agrupadas?		
11	Qual foi o tipo de vértice agrupado?	<i>Se necessário, marque mais de uma opção</i>	1) Exposições 2) Desfechos 3) Mediadores 4) Ancestrais de E e D 5) Outros
12	Algum DAG apresenta aspectos do desenho de estudo?		1) Sim 2) Não 3) Incerto
13	Que tipo de aspecto do desenho?	<i>Se necessário, marque mais de uma opção. Especifique o tipo de vértice</i>	1) Vértices indexados no tempo 2) Variáveis não observadas 3) Outros aspectos de mensuração 4) Aspectos de

			seleção 5) Outros
14	Qual o número de vértices do gráfico?	<i>No caso de múltiplos DAG, considere o gráfico com maior número de vértices</i>	
15	Qual o número de arestas do gráfico?	<i>No caso de múltiplos DAG, considere o gráfico com maior número de vértices</i>	

APÊNDICE C - Tesouro

Tesouro utilizado na padronização e exclusão de termos gerais para a construção dos mapas de coocorrência de palavras chaves nos títulos e resumos dos artigos. Na coluna da esquerda (label) são apresentados os termos de maior frequência no corpus e os seus substitutos a direita (coluna replace by). Foram excluídos das análises os termos com a ausência de substitutos na coluna replace by.

label	replace by
race/ethnicity	race
race ethnicity	race
ethnicity	race
white	race
black	race
african american	race
sex	gender
ses	socioeconomic factor
socioeconomic status	socioeconomic factor
active antiretroviral therapy	hiv
antiretroviral therapy	hiv
art	hiv
therapy	treatment
hiv-aids	hiv
hiv aids	hiv
hiv-infection	hiv
hiv infection	hiv
human-immunodeficiency virus	hiv
human immunodeficiency virus	hiv
immunodeficiency virus	hiv

environmental exposure	pollutants
perfluoroalkyl	pollutants
pfa	pollutants
pfas	pollutants
air pollution	pollutants
air pollution exposure	pollutants
air-pollution	pollutants
chemical	pollutants
chemicals	pollutants
particulate matter	pollutants
traffic pollution	pollutants
traffic	pollutants
phthalates	pollutants
phthalate	pollutants
bpa	pollutants
bisphenol a	pollutants
bisphenol	pollutants
pfos	pollutants
perfluorinated chemicals	pollutants
perfluorooctane Sulfonate	pollutants
perfluorinated chemicals	pollutants
pfhxs	pollutants
perfluorohexane sulfonic acid	pollutants
polybrominated diphenyl ethers	pollutants
pbdes	pollutants
pbde	pollutants
pfoa	pollutants
perfluorooctanoic-acid	pollutants
perfluorooctanoic acid	pollutants
polychlorinated biphenyls	pollutants

polychlorinated biphenyl	pollutants
pcb	pollutants
pcbs	pollutants
serum concentration	serum
blood	serum
plasma concentration	serum
blood sample	serum
sga	small for gestational age
small-for gestational age	small for gestational age
gwg	gestational weight gain
live birth	birth
preterm birth	birth
term birth	birth
birth outcome	birth
stillbirth	birth
child birth	birth
childbirth	birth
birth-weight	birth weight
low-birth-weight	birth weight
low birth weight	birth weight
lbw	birth weight
gestational	pregnancy
gestation	pregnancy
pregnant	pregnancy
prenatal exposure	pregnancy
late pregnancy	pregnancy
clinical pregnancy	pregnancy
early pregnancy	pregnancy
weeks gestation	pregnancy
pregnant woman	pregnancy

crp	c-reactive protein
bmi	obesity
adiposity	obesity
body-mass-index	obesity
body mass index	obesity
prepregnancy body mass index	obesity
obese	obesity
obese woman	obesity
overweight	obesity
waist	abdominal obesity
waist circumference	abdominal obesity
maternal obesity	obesity
hip ratio	abdominal obesity
fruit	fruit-vegetable
fruits	fruit-vegetable
vegetable	fruit-vegetable
vegetables	fruit-vegetable
diet	nutrition
food	nutrition
intake	nutrition
food frequency questionnaire	nutrition
dietary pattern	nutrition
dietary intake	nutrition
alcohol intake	alcohol
alcohol use	alcohol
alcohol consumption	alcohol
hydroxyvitamin d	vitamin d
25-Hydroxyvitamin	vitamin d
vitamin D supplementation	vitamin d
infertility	fertility

fertility treatment	fertility
vitro fertilization	fertility
vital status	mortality
survival	mortality
cancer mortality	mortality
cause mortality	mortality
death	mortality
breast cancer	cancer
breast cancer incidence	cancer
breast cancer risk	cancer
breast-cancer	cancer
cancers	cancer
malignant neoplastic disease	cancer
carcinoma	cancer
smoking status	smoking
smoker	smoking
current smoker	smoking
cigarette smoking	smoking
postmenopausal woman	menopause
premenopausal woman	menopause
depressive symptom	depression
kid	child
kids	child
children	child
young-child	child
young child	child
young-children	child
young children	child
infant	child
infants	child

infancy	child
early childhood	child
childhood	child
boy	child
girl	child
girls	child
boy	child
boys	child
offspring	child
toddler	child
mother child pair	child
man	men
male	men
female	women
woman	women
older woman	women
cvd	cardiovascular disease
cardiovascular-disease	cardiovascular disease
stroke	cardiovascular disease
atrial fibrillation	cardiovascular disease
heart failure	cardiovascular disease
myocardial infarction	cardiovascular disease
atherosclerosis	cardiovascular disease
sbp	blood pressure
systolic blood pressure	blood pressure
dbp	blood pressure
diastolic blood pressure	blood pressure
hypertension	blood pressure
diabete	diabetes
diabetes mellitus	diabetes

dag	
egrf	
low level	
etiology	
region	
literature	
evaluation	
present study	
deviation	
fold	
mean	
median	
proportion	
extent	
higher level	
degree	
context	
scale	
function	
multivariate analysis	
multivariable analysis	
study period	
adjusted model	
goal	
account	
hazard	
hazard ratio	
patient	
participant	
relation	

versus	
chinese china	
usa	
brazil	
msm multiple imputation	
propensity score inverse probability	
multivariate logistic regression unconditional logistic regression linear regression model	
linear regression models linear regressions	
cox model	
cox regression model	
cox proportional hazard model cox proportional hazards model	
confounders confounder	
potential confounder confounding	
cross-sectional study	cross sectional study
case-control case-control study	case control study case control study
prospective cohort study	logitudinal study
prospective cohort	logitudinal study
cohort	logitudinal study
retrospective cohort study	logitudinal study
birth cohort study	logitudinal study

birth cohort	logitudinal study
child cohort study	logitudinal study
longitudinal analysis	logitudinal study
cohort study	logitudinal study
person year	logitudinal
consequence	
mediation analysis	mediation
mediator	mediation
mechanism	mediation
causal pathway	mediation
directed effect	mediation
direct effect	mediation
indirected effect	mediation
indirect effect	Mediation

APÊNDICE D - Lista de pistas *Hedges*Pistas Hedges

may, almost, apparent , apparently, appear*, argue*, assume*, assumption, basically can, certain extent , conceivably, conclude*, conjecture*, consistent with, contention, could, could not, of course, deduce*, discern*, doubt*, essentially, establish* , expect*, the fact that, formally, frequently, general, generally, given that, guess*, however, hypothesize*, hypothetically, ideally, implication*, imply, improbable, indeed, indicate*, inevitable, interpret, we know, it is known, largely, likely, mainly, maybe, might, more or less, most, not necessarily, no doubt, beyond doubt, normally, occasionally, ostensibly, partially, partly, patently, perhaps, plausible, possibility, possible, possibly, postulate*, precisely, predict*, prediction, predominately, presumably, presume*, probability, probable, probably, propose*, prove*, provided that, open to question, questionable, quite, rare, rarely, rather, relatively, reportedly, reputedly, seem*, seems, seemingly, can be seen , seldom , general sense, should, speculate*, suggest, suggested, suggesting, suggestive, suppose*, surmise, suspect*, technically, tend*, tendency, theoretically, we think, typically, uncertain, unclear, unlikely, unsure, usually, virtually, would, would not

APÊNDICE E – Script em R**#### packages ####**

```

if(!require(rAltmetric))install.packages("rAltmetric");library(rAltmetric)
if(!require(bibliometrix))install.packages("bibliometrix");library(bibliometrix)
if(!require(RISmed))install.packages("RISmed");library(RISmed)
if(!require(plyr))install.packages("plyr");library(plyr)
if(!require(dplyr))install.packages("dplyr");library(dplyr)
if(!require(Rmisc))install.packages("Rmisc");library(Rmisc)
if(!require(scater))install.packages("scater");library(scater)
if(!require(reshape))install.packages("reshape");library(reshape)
if(!require(ggplot2))install.packages("ggplot2");library(ggplot2)
if(!require(readstata13))install.packages("readstata13");library(readstata13)
if(!require(quanteda))install.packages("quanteda");library(quanteda)
if(!require(tm))install.packages("tm");library(tm)
if(!require(logbin))install.packages("logbin");library(logbin)

```

```
#####
```

ANÁLISES BIBLIOMÉTRICAS

```
#####
```

#importar banco de dados do WOS/Pubmed

```
D1= readFiles('WOS_FINAL1.bib')
```

```
D2= readFiles('WOS_FINAL2.bib')
```

```
D3= readFiles('WOS_FINAL3.bib')
```

```
bibli1= convert2df(D1, dbsource = "isi", format = "bibtex")
```

```
bibli2= convert2df(D2, dbsource = "isi", format = "bibtex")
```

```
bibli3= convert2df(D3, dbsource = "isi", format = "bibtex")
```

```
bibliM=rbind.data.frame(bibli1 ,bibli2, bibli3)
```

```
#####
```

```
results <- biblioAnalysis(BibliM, sep = ";")
```

```
summary(results, k=30)
```

####Extração dos dados

```
#País de afiliação
```

```
BibliM = metaTagExtraction(BibliM, Field = "AU_CO", sep = ";")
```

#País primeiro autor

```
BibliM= metaTagExtraction(BibliM, Field = "AU1_CO", sep = ";")
```

#Frequência de países de afiliação do primeiro autor

```
A1 = cocMatrix(BibliM, Field = "AU_CO", sep = ";")
```

```
sort(Matrix::colSums(A), decreasing = TRUE)[1:86]
```

#coautores

```
Ac = cocMatrix(BibliM, Field = "AU1_CO", sep = ";")
```

```
sort(Matrix::colSums(A), decreasing = TRUE)[1:85]
```

#Afiliação

```
BibliM= metaTagExtraction(BibliM, Field = "AU_UN", sep = ";")
```

#Temas

```
BibliM= metaTagExtraction(BibliM, Field = "DE", sep = ";")
```

```
#####3
```

#Comparar Rede de coocorrência sem o tesauros

```
Banco_keyword=read.csv("Banco_keywords.csv", header = T)
```

```
Banco_keyword[] = lapply(Banco_keyword, function(x) gsub("EPIDEMIOLOGIC  
METHODS", "", x))
```

```
Banco_keyword[] <- lapply(Banco_keyword, function(x) gsub("RISK", "", x))
```

```

banco_bibliDAG$ID=Banco_keyword$authors_keywords
NetMatrix = biblioNetwork(banco_bibliDAG, analysis = "co-occurrences", network =
"keywords", sep = ";")
net=networkPlot(NetMatrix, normalize="association", weighted=T, n = 40,
Title = "Co-ocorrência de palavras chaves", type = "vosviewer", size=T, edgesize =
5,labels=0.8)

```

```
#####Número de autores#####
```

```
#Verificar erros, comparação contagem por dados de afiliação/nome
```

```
sep_autores= lapply(dados_totais_bibliometrica2$AU, count.fields(textConnection(),
sep = ";"))
```

```
dados_totais_bibliometrica2$sep_autores_AU =
sapply(dados_totais_bibliometrica2$AU, function(x)
length(unlist(strsplit(as.character(x), ";"))))
```

```
dados_totais_bibliometrica2$sep_autores_country =
sapply(dados_totais_bibliometrica2$AU_CO,
function(x) length(unlist(strsplit(as.character(x), ";"))))
```

```
summary(dados_totais_bibliometrica2$sep_autores)
numero_autores= apply(dados_totais_bibliometrica2$sep_autores, 2, sum)
sum(dados_totais_bibliometrica2$sep_autores)
```

```
#####
```

```
#Gráfico número de publicações por ano
```

```
#####
```

```
#Preparação dos dados
```

```
data_graph=read.csv("Contagem_publicacoes(GERAL e DAG).csv", header = T)
data_graph$DAG_norm <- (data_graph$DAG / data_graph$TOTAL_PUBMED*100)
data_graph$DAG_Etio_norm <-(data_graph$DAG / data_graph$Etiology_key*100)
data_graph$norm_etiology= (data_graph$Etiology_key /
data_graph$TOTAL_PUBMED*100)
```

```

data_graph$norm_title_causality= (data_graph$Busca_tituloResumo.PUBMED./
data_graph$TOTAL_PUBMED*100)
write.csv(data_graph, "Contagem_publicacoes(GERAL e DAG)2.csv")

data_graph=read.csv("Contagem_publicacoes(GERAL e DAG)2.csv", header = T)
data_graph=data_graph[,c(-13,-11,-12)]
colnames(data_graph)[11] <- "DAG/Total de artigos indexados"
colnames(data_graph)[12] <- "DAG/Etiologia"
colnames(data_graph)[13] <- "Etiologia/Total"
colnames(data_graph)[14] <- "Causalidade(Título e resumo)/Total"
data_graph_melt=data_graph[,c(2,11,12)]
data_graph_melt_cum=data_graph[,c(2,13)]

# transform the data - melt
df <- melt(data_graph_melt, id = c("year"))
colnames(df)[2] <- "Comparações"

#####
#Gráficos publicações por ano
#####
#Gráfico 1A
theme_update(plot.title = element_text(hjust = 0))

plot2=ggplot(df, aes(x =year, y=value,
                    group = Comparações))+
  geom_line(aes(linetype=Comparações))+
  scale_x_continuous(expand = c(0,0), breaks =
c(2000,2002,2004,2006,2008,2010,2012,2014,2016))+
  ggtitle("Número relativo de publicações", subtitle="Base de dados: Medline (2000-
2017)") +
  xlab("Ano") +
  ylab("%")+

```



```

theme(axis.title.y = element_text(family = "sans", color="black", size=10,
face="bold"),
      plot.title = element_text(family = "sans", size=13, face="bold", hjust = 0),
      plot.subtitle=element_text(size=10, hjust=0, face="italic", color="black"),
      panel.border = element_rect(linetype = "solid", fill = NA),
      legend.position="bottom",
      legend.title = element_blank())

```

#Gráfico 1B

```

plot1= ggplot(data_graph, aes(x = year, y=DAG))+
geom_bar(stat="identity", width=0.95, fill="black") +
  scale_x_continuous(expand = c(0,0), breaks =
c(2000,2002,2004,2006,2008,2010,2012,2014,2016))+
  ggtitle("Publicações com a utilização de diagramas causais",
        subtitle="Base de dados: Medline, WoS, PMC - (2000-2017)") +
  xlab("Ano") +
  ylab("Número de artigos")+
  theme(axis.title.y = element_text(family = "sans", color="black", size=10, hjust = 0.5,
vjust = 1),
        plot.title = element_text(family = "sans", size=12, face="bold",hjust = 0),
        plot.subtitle=element_text(size=10, hjust=0, face="italic", color="black"),
        panel.border = element_rect(linetype = "solid", fill = NA))+
  geom_text(aes(label = DAG), position = position_dodge(width = 1), vjust = -0.5, size
= 3)

require(cowplot)
plot_final=plot_grid(plot1, plot2, labels = "AUTO", ncol = 1, align = "v")
plot_final
# Run multiplot function - end of script
multiplot(plot1, plot2, cols =1)

```

```
#####
```

#Gráfico fontes de publicação

#####

#preparacao dados

```

banco_bibliDAG=read.csv("Banco_bibliFDAG.csv", header =T)
journal=as.character(banco_bibliDAG$JI)
banco_bibliDAG$journal3=sapply(as.character(banco_bibliDAG$journal2),
removePunctuation)
table(banco_bibliDAG$journal3)
tab_journal_s=sort(table(banco_bibliDAG$journal3), decreasing=TRUE)
tab_journal1=sort(tab_journal[tab_journal>=7], decreasing=F)
articles=as.numeric(tab_journal1,2)
group= ifelse(articles==81,1, ifelse(articles==58,2, ifelse(articles==40,3,
ifelse(articles==22,4,
ifelse(articles==21,5, ifelse(articles==19, 6, ifelse(articles==18,7,
ifelse(articles==17,8,
ifelse(articles==14,9, ifelse(articles==13,10, ifelse(articles==12,11,
ifelse(articles==11,12,
ifelse(articles==9,13,ifelse(articles==8,14, ifelse(articles==7,15,NA)))))))))))))
levels1=names(tab_journal1)
dat <- data.frame(articles, levels1, group)
write.csv(dat, file="revista.csv")

```

#coletar dados de fator de impacto 5 anos JCR

#####

#Gráfico principais revistas de publicacao

#####

#banco com dados do FI JCR

```

revistas=read.csv("revista2.csv")
dat=revistas[,c(2,3,4,7)]
colnames(dat)[4] <-"FI_JCR"

```

#Gráfico 3

```

p= ggplot(data=dat, aes(reorder(levels1, articles), articles, fill=FI_JCR))
p + geom_bar(stat="identity", width=.7) +
  coord_flip() +
  scale_fill_manual(values=c("darkgray", "dimgray", "black"))+
  scale_y_continuous(expand = c(0,4), breaks = c())+
  ggtitle("Principais revistas de publicação", subtitle="Base de dados: Scopus, WoS,
PMC - (2000-2017)") +
  xlab("") +
  ylab("Número de artigos com o relato da utilização de diagramas causais")+
  geom_text(aes(label = articles), hjust = -0.5, size = 3, position =
position_dodge(width = 1), inherit.aes = TRUE)+
  theme(axis.title.y = element_text(family = "sans", color="black", size=5, hjust = 0,
vjust = 1),
        axis.title.x = element_text(family = "sans", color="black", size=10, hjust = 0, vjust
= 1),
        plot.title = element_text(family = "sans", size=14, face="bold"),
        panel.border = element_rect(fill = NA))

```

```
#####
```

Altmetrics coleta de dados

```
#####
```

Importar banco com doi_csv file with ref_id / doi/

```

refs=read.csv("dados finais para analise bibliometrica.csv", header=T,
stringsAsFactors = T)

```

```
refs <- subset(refs,doi!="")
```

Primeiro, criar lista de doi.

```

doi_list2= as.list(strsplit(as.character(refs$doi), ","))
novos_pmid=read.csv("novos_pmid.csv", header = T)
novos_pmid$pmid_litC= noquote(paste0(novos_pmid$pmid_C, ","))
novos_pmid$pmid_litD= noquote(paste0(novos_pmid$pmid_D, ","))
write.csv(novos_pmid, file="novos_pmid2.csv")

```

Segundo, coleta dos dados de altimetria (1034 artigos = ~50min)

```
metric_data_final =ldply(doi_list2, function(x)
tryCatch(altmetric_data(altmetrics(doi = x)), error=function(e) NULL,
          silent=T), .progress = 'text')

write.csv(metric_data_final, file = "metric_data_marco2018.csv")
```

#####Analise altmetrics

```
metric_data2= read.csv2("metric_data_marco2018.csv", header=T, sep="," ,
stringsAsFactors = F)
dados_totais_bibliometrica2=read.csv("dados_totais_bibliometrica2.csv", header=T,
stringsAsFactors = T)
```

#Reduzir tamanho do banco

```
metric_data_menor=metric_data2[,c("url", "score", "context.all.mean",
"context.all.rank", "context.all.pct",
          "context.journal.pct", "context.similar_age_3m.pct",
context.similar_age_journal_3m.pct",
          'cited_by_tweeters_count', 'cited_by_gplus_count',
cited_by_fbwalls_count",
          "cited_by_feeds_count")]
```

#Escore altmetrics distrib.

```
summary.data.frame(banco_alt_new)
banco_alt_new$score=as.numeric(banco_alt_new$score)
summary(banco_alt_new$score)
banco_alt_new$score2=banco_alt_new$score
banco_alt_new$score2[is.na(banco_alt_new$score2)]= 0
tapply(banco_alt_new$score, banco_alt_new$status, summary)
tapply(banco_alt_new$score2, banco_alt_new$status, summary)
tapply(banco_alt_new$TC, banco_alt_new$status, summary)
```

```
tapply(banco_alt_new$TC, banco_alt_new$status, summary)
banco_alt_new$TC=as.numeric(banco_alt_new$TC)
```

#Principais tipos de citações

```
summary(banco_alt_new$cited_by_tweeters_count)
summary(banco_alt_new$cited_by_gplus_count)
summary(banco_alt_new$cited_by_fbwalls_count)
summary(banco_alt_new$cited_by_feeds_count)
summary(banco_alt_new$cited_by_feeds_count)
```

#Altmetrics em contexto

#Rank

```
context_rank=as.numeric(banco_alt_new$context.all.rank)
summary (context_rank)
```

#Percentil geral

```
context_pct=as.numeric(banco_alt_new$context.all.pct)
summary(context_pct)
```

#Número de publicações no percentil 80 ou acima

```
context_pct2 <- ifelse(banco_alt_new$context.all.pct<79, "<PCT80", ">=PCT80")
contextJ_pct=as.numeric(banco_alt_new$context.journal.pct)
summary (contextJ_pct)
contextJ_pct2 <- ifelse(banco_alt_new$context.journal.pct<79, "<PCT80",
">=PCT80")
table(contextJ_pct2)
contextA_pct=as.numeric(banco_alt_new$context.similar_age_3m.pct)
summary (contextA_pct)
contextA_pct2 <- ifelse(banco_alt_new$context.similar_age_3m.pct<79, "<PCT80",
">=PCT80")
table(contextA_pct2)
contextJA_pct=as.numeric(banco_alt_new$context.similar_age_journal_3m.pct)
summary (contextJA_pct)
```

```
contextJA_pct <- ifelse(banco_alt_new$context.similar_age_journal_3m.pct<79,
"<PCT80", ">=PCT80")
table(contextJA_pct)
```

Recuperar dados faltantes

```
missing_alt_dados=subset(banco_alt_new, score==0)
# First, create a list of dois.
missing_alt_dados <- subset(missing_alt_dados, doi.x!="")
doi_list3= as.list(strsplit(as.character(missing_alt_dados$doi.x), ","))
paste(missing_WOS$doi.x, "OR", sep = ", OR")
write.csv(missing_WOS, file="missing_wos.csv")
metric_data_missing$doi=sapply(dados_pre_bibliometrica$doi, toupper)
doi_list3=lapply(doi_list3, tolower)
```

##Comparação com DAG no artigo e sem DAG no artigo

```
com_DAGartigo=read.csv("artigos_comdag_totalNEW.csv")
colnames(com_DAGartigo) <- c("record")
com_DAGartigo$status="Com DAG artigo"
banco_bibli_new=join(banco_bibliDAG, com_DAGartigo, by="record")
banco_bibli_new$status[is.na(banco_bibli_new$status)]= "Sem DAG artigo"

banco_metric_new=join(banco_bibliDAG, com_DAGartigo, by="record")
banco_bibli_new$status[is.na(banco_bibli_new$status)]= "Sem DAG artigo"

banco_paramerge=banco_bibli_new[,c("record", "doi", "AU", "journal", "status", "TC")]
metric_data_merge=metric_data_final[,c("doi", "score", "authors", "context.all.mean",
"context.all.rank", "context.all.pct",
"context.journal.pct", "context.similar_age_3m.pct",
"context.similar_age_journal_3m.pct",
'cited_by_tweeters_count', 'cited_by_gplus_count',
" cited_by_fbwalls_count", "cited_by_feeds_count",
```

```

      "cited_by_peer_review_sites_count",
      "cited_by_policies_count")]
```

```

metric_data_merge$doi=sapply(metric_data_merge$doi, toupper)
banco_alt_new=join(banco_paramerge, metric_data_merge, by="doi")
```

#Escore altmetrics distrib dados recuperados.

```

banco_alt_new$score=as.numeric(banco_alt_new$score)
summary(banco_alt_new$score)
banco_alt_new$score2=banco_alt_new$score
banco_alt_new$score2[is.na(banco_alt_new$score2)]= 0
```

```

tapply(banco_alt_new$score, banco_alt_new$status, summary)
tapply(banco_alt_new$score2, banco_alt_new$status, summary)
tapply(banco_alt_new$TC, banco_alt_new$status, summary)
```

```

y1=subset(banco_alt_new, banco_alt_new$status=="Com DAG artigo")
y2=subset(banco_alt_new, banco_alt_new$status=="Sem DAG artigo")
```

```

banco_citation_semOUT=subset(banco_alt_new, banco_alt_new$TC<100)
y22=subset(y2, y2$TC<100)
t.test(y2$score2, y1$score2, alternative = "two.sided", var.equal = FALSE)
t.test(y11$TC, y22$TC, alternative = "two.sided", var.equal = FALSE)
t.test(citations ~ status, data = banco_alt_new, alternative = "two.sided")
```

histogram contagem de citações

```

hist(y11$TC, col="grey")
```

```

plot(density(y11$TC), bty="l")
abline(v=quantile(y11$TC, c(0.5, 0.95)), lty=2:3)
require(cowplot)
plot_final=plot_grid(plot1, plot2, labels = "AUTO", ncol = 1, align = "v")
```

```

multiplot(plot1, plot2, cols =1)
save_plot("plot_ano.tiff", plot_final)
plot_final
#####
Gráfico Escore de altmetria e contexto/citações
#####
#altmetria
alt_context_new <- rbind(data.frame(Contexto="Geral",
obs=as.numeric(banco_alt_new$context.all.pct)),
                        data.frame(Contexto="Revista",
obs=as.numeric(banco_alt_new$context.journal.pct)),
                        data.frame(Contexto="Idade",
obs=as.numeric(banco_alt_new$context.similar_age_3m.pct)),
                        data.frame(Contexto="Revista e Idade",
obs=as.numeric(banco_alt_new$context.similar_age_journal_3m.pct)))
#altmetrics
ggplot(alt_context_new, aes(x=obs, y=obs, group= Contexto)) +
  geom_line() +
  ggtitle("Altimetric Attention Score", subtitle="Contexto do escore") +
  ylab("Percentil")

hist(alt_context_new$obs)
geom_count(as.numeric(banco_alt_new$context.all.pct))

citations_new <- rbind(data.frame(status="Sim", obs=as.numeric(y11$TC)),
                      data.frame(status="Não", obs=as.numeric(y22$TC)))
#Citations
ggplot(citations_new, aes(x= status, y=obs)) +
  geom_boxplot(na.rm = T) +
  ggtitle("Citações", subtitle="") +
  ylab("Número")

```



```
#####
CARACTERÍSTICAS DOS MODELOS GRÁFICOS E O RELATO DE
CONFUNDIMENTO
#####
```

```
banco_DAG_FINAL=read.dta13('DAG_FINAL(100).dta')
colnames(banco_DAG_FINAL) = c('date', 'avaliador', 'PMID', 'record' (.....)
```

```
# Análises descritivas modelos graficos
```

```
lapply(banco_DAG_FINAL[,c('dag_artigo', 'dag_referencial', 'dag_incerteza',
'auseprese_variaveis', 'auseprese_variaveis_Sm', 'auseprese_setas',
      'auseprese_setas_Sm', 'direcao_setas',
      'direcao_setas_Sm', 'ciclos', 'ciclos_Sm',
      'papel_variavel', 'papel_variavel_Sm', 'outra_incerteza',
      'dag_utilizacao', '1DAG_multiplas_analises',
      '1DAG_multiplas_analises_motiv',
      'multiplosDAG', 'variaveis_agrupadas',
      'DAG_desenho', 'indexacao_tempo',
      'variaveis_naoobser', 'mensuracao',
      'selecao', 'outro_aspecto')], function(x) data.frame( table(x),
prop.table( table(x))))
```

```
#Densidade dos DAG
```

```
lapply(banco_DAG_FINAL[,c("dag_numero", 'numero_vertices', 'numero_arestas',
'variaveis_agrupadas_n')], summary)
```

```
banco_DAG_FINAL$maximal_Edge_count= ((banco_DAG_FINAL$numero_vertices-
1)*(banco_DAG_FINAL$numero_vertices))/2
banco_DAG_FINAL$density=banco_DAG_FINAL$numero_arestas/banco_DAG_FIN
AL$maximal_Edge_count
summary(banco_DAG_FINAL$density)
table(banco_DAG_FINAL$density)
```

```
banco_graph=banco_DAG_FINAL[c("record", "numero_vertices", "numero_arestas",
"maximal_Edge_count", "density")]
```

Gráfico variáveis agrupadas

```
n_amais2=(banco_DAG_FINAL$n_variaveis)-(banco_DAG_FINAL$numero_vertices
-2)
a=hist(n_amais2)
```

```
ggplot(banco_graph, aes(n_amais2)) +
geom_histogram(breaks= c(-15,-10,-5, 0, 5, 10,15), colour="black", fill="white" )+
  scale_x_continuous(breaks= c(-15,-10,-5, 0, 5, 10,15), labels=format(c("-15","-10","-
5", "0","5","10","15")))+
  theme(axis.title.x = element_text(family = "sans", color="black", size=14, hjust =
0.5, vjust = 1), plot.title = element_text(family = "sans", size=16, face="bold"))+
ggtitle("Características dos diagramas causais", subtitle="Diferença entre o número
de confundidores ajustados e o número de vértices do DAG") +
  xlab("Diferença") +
  ylab("Frequência")
```

```
#####
```

Análises descritivas relato de confundimento

```
#####
```

```
lapply(banco_DAG_FINAL[,c('relato_razoes_PC', 'estrategia_PC',
'relato_razoes_modelofinal', 'criterios_modelofinal', 'table2fallacy',
      'table2_interpretacao', 'relato_metodoajuste',
      'metodo_ajuste', 'relato_anal_sensibilidade',
'relato_anal_sensibilidade_tradicional',
      'modelagem_vies', 'relato_estimativas', 'relato_caracteriscas',
'dag_utilizacao', '1DAG_multiplas_analises',
'1DAG_multiplas_analises_motiv',
'multiplosDAG', 'variaveis_agrupadas',
'DAG_desenho', 'indexacao_tempo',
```

```
'variaveis_ naoobser', 'mensuracao',
'selecao', 'outro_aspecto')], function(x) data.frame( table(x),
prop.table( table(x))))
```

```
table(banco_DAG_FINAL$estrategia_o,
banco_DAG_FINAL$estrategia_PC=="Outros")
table(banco_DAG_FINAL$criterioMF_o,
banco_DAG_FINAL$criterios_modelofinal=="Outro")
table(banco_DAG_FINAL$table2_coment)
table(banco_DAG_FINAL$outro_metodoajuste1)
table(banco_DAG_FINAL$outro_metodoajuste2)
table(banco_DAG_FINAL$tradicional_coment)
table(banco_DAG_FINAL$modelagem_vies_coment)
table(banco_DAG_FINAL$comentarios_confundimento,
banco_DAG_FINAL$relato_estimativas=="Incerto (especificar)")
table(banco_DAG_FINAL$comentarios_confundimento,
banco_DAG_FINAL$relato_caracteriscas=="Incerto (especificar)")
summary(banco_DAG_FINAL$n_variaveis)
```

```
#####
```

Avaliação de confundimento não observado

```
#####
```

```
limitacoes= banco_DAG_FINAL[,c('tipo_limitacao1','tipo_limitacao2',
'tipo_limitacao3', 'tipo_limitacao4', 'tipo_limitacao5', 'tipo_limitacao6',
'tipo_limitacao7')]
```

```
limitacoes$Confundimento=apply(limitacoes,1,function(x) {sum(x =='Confundimento
nao observado/residual' , na.rm = TRUE)})
```

```
limitacoes$Confundimento_mediacao=apply(limitacoes,1,function(x) {sum(x
=='Confundimento entre o mediador e o desfecho' , na.rm = TRUE)})
```

```
limitacoes$Confundimento_intermediario=apply(limitacoes,1,function(x) {sum(x
=='Confundimento intermediario' , na.rm = TRUE)}))
```

#Reconhecimento de confundimento

```
lapply(limitacoes[,c('Confundimento', 'Confundimento_mediacao',
'Confundimento_intermediario')], function(x) data.frame( table(x),
prop.table( table(x))))
```

##Avaliação da direção do viés de confundimento

```
record1=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao1=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao1=="Sim"),
"record"]
```

```
record2=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao2=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao2=="Sim"),
"record"]
```

```
record3=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao3=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao3=="Sim"),
"record"]
```

```
record4=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao4=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao4=="Sim"),
"record"]
```

```
record5=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao5=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao5=="Sim"),
"record"]
```

```
record6=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao6=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao6=="Sim"),
"record"]
```

```
record7=banco_DAG_FINAL[which(banco_DAG_FINAL$tipo_limitacao7=="Confundi
mento nao observado/residual" & banco_DAG_FINAL$avalicao_direcao7=="Sim"),
"record"]
```

```
confundimento_records_d=(c(record1, record2, record3, record4, record5, record6,
record7))
```

```
length(unique(c(record1, record2, record3, record4, record5, record6, record7)))
```

```
#####
```

RELATO DAS LIMITAÇÕES DO ESTUDO

```
#####
```

#Termos hedges

#Preparação do corpus

```
# pdftotext
```

```
#cd /home/taisacortes/....
```

```
#for file in *.pdf; do pdftotext "$file" "$file.txt"; done
```

```
#edição no getid
```

```
#extração da discussão, respeitando o layout
```

```
#####
```

```
corpus_DAG <-
```

```
Corpus(DirSource(directory="/home/taisacortes/Documentos/ARTIGOS/DAG(FINAL  
100ARTIGOS)/TXT(100)"))
```

```
corpus_controle <-
```

```
Corpus(DirSource(directory="/home/taisacortes/Documentos/ARTIGOS/CONTROLE  
(FINAL100ARTIGOS)/TXT(100)"))
```

```
corpus.p <-tm_map(corpus_DAG, removeNumbers)
```

```
corpus.p <-tm_map(corpus.p, removePunctuation)
```

```
dtm <-DocumentTermMatrix(corpus.p)
```

```
median(rowSums(as.matrix(dtm)))
```

```
#####
```

#Summary do corpus, contagem de palavras (tokens, sentencas)

```
mycorpus <- corpus(corpus_controle)
```

```
corpus_controle_NEW=summary(mycorpus)
```

```
write.csv(corpus_controle_NEW, file = "corpus_CONc_NEW.csv")
```

```
write.csv(corpus_DAG_NEW, file="corpus_dag_NEW.csv")
```

#Upload do banco

```
corpus_DAG_N=read.csv("corpus_dag_final.csv", header = T)
```

```

corpus_DAG_N$status="Com DAG"
corpusC_N=read.csv("corpus_CON_Final2.csv", header = T)
corpusC_N$status="Sem DAG"
corpus_todos=rbind(corpus_DAG_N, corpusC_N)

```

#Hedges normalizado = resultado * (tamanho desejado(100) / numero de tokens no corpus)

```

corpus_todos$rowS=corpus_todos$hedges /corpus_todos$Tokens
corpus_todos$rowSN=corpus_todos$hedges*(100/corpus_todos$Tokens)

```

#Análise descritiva

```

tapply(corpus_todos$hedges, corpus_todos$status, summary)
tapply(corpus_todos$Tokens, corpus_todos$status, summary)
tapply(corpus_todos$Sentences, corpus_todos$status, summary)
tapply(corpus_todos$rowSN, corpus_todos$status, summary)

```

```

t.test(corpus_todos$rowSN ~ corpus_todos$status, alternative = "two.sided",
var.equal = FALSE)

```

##Gênero e etnia primeiro autor

```

genni_DAG=read.csv("Ethnea_TODOS.csv", header = T)
genni_C=read.csv("Ethnea_novosC.csv", header = T)
genni_DAG$status="Com DAG"
genni_C$status="Sem DAG"
genni_final=rbind(genni_DAG, genni_C)
genni_final$Ethnea2=ifelse(genni_final$Ethnea=="ENGLISH", "inglesa", "nao
inglesa")
table(genni_final$Ethnea2, genni_final$status)
table(genni_final$Genni, genni_final$status)
tapply(genni_final$n_autores, genni_final$status, summary)

```

```
#####
```

#Relato limitações

#banco métodos

```
DAG_FINAL=read.csv("banco_DAG_FINAL.cvs", header = T)
DAG_FINAL=banco_DAG_FINAL[,c(1:94)]
DAG_FINAL$status="Com DAG"
controle_FINAL=read.csv("controle_final.csv", header = T)
controle_FINAL$status="Sem DAG"
metodos_d= DAG_FINAL[,c('status', 'avaliador', 'PMID', 'record', 'desenho',
'desenho_o', 'objetivo1', 'objetivo1_o',
'objetivo2', 'objetivo2_o', 'exposicao_tipo', 'exposicao_V1', 'exposicao_V2',
'desfecho_tipo', 'desfecho_V1',
'desfecho_V2', 'outra_var_tipo', 'outra_var', 'outra_var_tipo2', 'outra_var2',
'local', 'n_amostra')]

metodos_c= controle_FINAL[,c('status', 'avaliador', 'PMID', 'record', 'desenho',
'desenho_o', 'objetivo1', 'objetivo1_o',
'objetivo2', 'objetivo2_o', 'exposicao_tipo', 'exposicao_V1',
'exposicao_V2', 'desfecho_tipo', 'desfecho_V1',
'desfecho_V2', 'outra_var_tipo', 'outra_var', 'outra_var_tipo2',
'outra_var2', 'local', 'n_amostra')]

metodos_final=rbind(metodos_d, metodos_c)
table(metodos_final$desenho, metodos_final$status)
table(metodos_final$objetivo1, metodos_final$status)
table(metodos_d$objetivo1, metodos_d$objetivo2)
table(metodos_c$objetivo1, metodos_c$objetivo2)
tapply(metodos_final$n_amostra, metodos_final$status, summary)
```

```
#####Banco limitações
```

```
limitacoes_d= DAG_FINAL[,c('status', 'record', 'tipo_limitacao1','tipo_limitacao2',
'tipo_limitacao3', 'tipo_limitacao4', 'tipo_limitacao5', 'tipo_limitacao6', 'tipo_limitacao7',
'avalicao_direcao1', 'avalicao_direcao2',
      'avalicao_direcao3', 'avalicao_direcao4',
      'avalicao_direcao5', 'avalicao_direcao6',
      'avalicao_direcao7', 'avalicao_quantita1', 'avalicao_quantita2',
      'avalicao_quantita3', 'avalicao_quantita4',
      'avalicao_quantita5', 'avalicao_quantita6',
      'avalicao_quantita7')]
```

```
limitacoes_cc= controle_FINAL[,c('status','record', 'tipo_limitacao1','tipo_limitacao2',
'tipo_limitacao3', 'tipo_limitacao4', 'tipo_limitacao5', 'tipo_limitacao6', 'tipo_limitacao7',
'avalicao_direcao1', 'avalicao_direcao2', 'avalicao_direcao3', 'avalicao_direcao4',
'avalicao_direcao5', 'avalicao_direcao6', 'avalicao_direcao7', 'avalicao_quantita1',
'avalicao_quantita2', 'avalicao_quantita3', 'avalicao_quantita4', 'avalicao_quantita5',
'avalicao_quantita6',
      'avalicao_quantita7')]
```

```
banco_limitacoes_final=rbind(limitacoes_d, limitacoes_cc)
```

#Contagem por tipo de limitação

#Confundimento

```
banco_limitacoes_final$Confundimento=apply(banco_limitacoes_final,1,function(x)
{sum(x =='Confundimento nao observado/residual' , na.rm = TRUE)})
```

#Confundimento mediação

```
banco_limitacoes_final$Confundimento_mediacao=apply(banco_limitacoes_final,1,funct
nction(x) {sum(x =='Confundimento entre o mediador e o desfecho' , na.rm =
TRUE)})
```

#Outra variável não observada

```
banco_limitacoes_final$Outra_variavel_naoobs=apply(banco_limitacoes_final,1,funct
ion(x) {sum(x =='Outras variaveis nao observadas' , na.rm = TRUE)})
```


#Mensuração da exposição

```
banco_limitacoes_final$Mensuracao_Exposicao=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Erros de mensuracao da exposicao' , na.rm = TRUE)})
```

#Mensuração do desfecho

```
banco_limitacoes_final$Mensuracao_Desfecho=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Erros de mensuracao do desfecho' , na.rm = TRUE)})
```

#Mensuração do mediador

```
banco_limitacoes_final$Mensuracao_Mediador=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Erros de mensuracao do mediador ' , na.rm = TRUE)})
```

#Mensuração geral

```
banco_limitacoes_final$Mensuracao_geral=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Erros de mensuracao (geral)' , na.rm = TRUE)})
```

#Dados faltantes

```
banco_limitacoes_final$Dados_faltantes=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Dados faltantes' , na.rm = TRUE)})
```

#Perda de seguimento

```
banco_limitacoes_final$Perda_seguinto=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Perda de seguimento' , na.rm = TRUE)})
```

#Viés de seleção (geral)

```
banco_limitacoes_final$Vies_selecao=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Vies de selecao (geral)' , na.rm = TRUE)})
```

#Temporalidade

```
banco_limitacoes_final$Tempo=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Tempo' , na.rm = TRUE)})
```

#Validade externa/transportabilidade

```
banco_limitacoes_final$Validade_externa=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Validade externa' , na.rm = TRUE)})
```

#Tamanho da amostra

```
banco_limitacoes_final$Tamanho_amostra=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Tamanho da amostra' , na.rm = TRUE)})
```

```
banco_limitacoes_final$Outra=apply(banco_limitacoes_final,1,function(x) {sum(x == 'Outra' , na.rm = TRUE)})
```

#Verificar se há erro de contagem

```
banco_limitacoes_final$n_limitacoes=rowSums(!is.na(banco_limitacoes_final[,c('tipo
_limitacao1','tipo_limitacao2', 'tipo_limitacao3', 'tipo_limitacao4', 'tipo_limitacao5',
'tipo_limitacao6', 'tipo_limitacao7')]))
```

```
banco_limitacoes_final$n_limitacoes2=rowSums(banco_limitacoes_final[,c('Confundi
mento', 'Confundimento_mediacao', 'Outra_variavel_naoobs',
'Mensuracao_Exposicao', 'Mensuracao_Desfecho', 'Mensuracao_Mediador',
'Mensuracao_geral', 'Dados_faltantes', 'Perda_seguinto', 'Vies_selecao', 'Tempo',
'Validade_externa', 'Tamanho_amostra', 'Outra')])
```

```
identical(banco_limitacoes_final[['n_limitacoes']],banco_limitacoes_final[['n_limitacoe
s2']])
```

```
#OK
```

```
#####
```

Discussão qualitativa de viés

```
#####
```

```
banco_limitacoes_final$Discutiu=apply(banco_limitacoes_final,1,function(x) {sum(x
=='Sim' , na.rm = TRUE)})
```

```
banco_limitacoes_final$Quantifi_AS=apply(banco_limitacoes_final,1,function(x)
{sum(x=='Sim, analise de sensibilidade' , na.rm = TRUE)})
```

```
banco_limitacoes_final$Quantifi_MV=apply(banco_limitacoes_final,1,function(x)
{sum(x=='Sim, modelagem de vies' , na.rm = TRUE)})
```

#Apenas reconheceu as limitações do estudo

```
banco_limitacoes_final$Quantificou= banco_limitacoes_final$Quantifi_AS +
```

```
banco_limitacoes_final$Quantifi_MV
```

```
banco_limitacoes_final$Avaliou= banco_limitacoes_final$Discutiu +
```

```
banco_limitacoes_final$Quantifi_MV + banco_limitacoes_final$Quantifi_AS
```

banco_limitacoes_final\$Apenas_reconheceu=ifelse(banco_limitacoes_final\$Avaliou==0,"Sim","Nao")

banco_limitacoes_final\$Discutiu2=ifelse(banco_limitacoes_final\$Discutiu==0,"Nao","Sim")

banco_limitacoes_final\$Quantificou2=ifelse(banco_limitacoes_final\$Quantificou==0,"Nao","Sim")

#####

Categorização do tipo de vies

#####

banco_limitacoes_final\$Confundimento2=ifelse(banco_limitacoes_final\$Confundimento==0,"Nao","Sim")

banco_limitacoes_final\$Confundimento_mediacao2=ifelse(banco_limitacoes_final\$Confundimento_mediacao==0,"Nao","Sim")

banco_limitacoes_final\$Outra_variavel_naoobs2=ifelse(banco_limitacoes_final\$Outra_variavel_naoobs==0,"Nao","Sim")

banco_limitacoes_final\$Mensuracao_Exposicao2=ifelse(banco_limitacoes_final\$Mensuracao_Exposicao==0,"Nao","Sim")

banco_limitacoes_final\$Mensuracao_Desfecho2=ifelse(banco_limitacoes_final\$Mensuracao_Desfecho==0,"Nao","Sim")

banco_limitacoes_final\$Mensuracao_Mediador2=ifelse(banco_limitacoes_final\$Mensuracao_Mediador==0,"Nao","Sim")

banco_limitacoes_final\$Mensuracao_geral2=ifelse(banco_limitacoes_final\$Mensuracao_geral==0,"Nao","Sim")

banco_limitacoes_final\$Dados_faltantes2=ifelse(banco_limitacoes_final\$Dados_faltantes==0,"Nao","Sim")

banco_limitacoes_final\$Perda_seguimento2=ifelse(banco_limitacoes_final\$Perda_seguimento==0,"Nao","Sim")

banco_limitacoes_final\$Vies_selecao2=ifelse(banco_limitacoes_final\$Vies_selecao==0,"Nao","Sim")

```
banco_limitacoes_final$Tempo2=ifelse(banco_limitacoes_final$Tempo==0, "Nao",
"Sim")
```

```
banco_limitacoes_final$Validade_externa2=ifelse(banco_limitacoes_final$Validade_
externa==0, "Nao", "Sim")
```

```
banco_limitacoes_final$Tamanho_amostra2=ifelse(banco_limitacoes_final$Tamanho
_amostra==0, "Nao", "Sim")
```

```
banco_limitacoes_final$Outra2=ifelse(banco_limitacoes_final$Outra==0, "Nao",
"Sim")
```

#Mensuração (geral)

```
banco_limitacoes_final$Mensuracao_geral3 <- with(banco_limitacoes_final,
ifelse(Mensuracao_geral2=="Sim" |Mensuracao_Desfecho2=="Sim" |
Mensuracao_Exposicao2=="Sim" | Mensuracao_Mediador2=="Sim", 1, 0))
```

#Confundimento (geral)

```
banco_limitacoes_final$Confundimento3 <- with(banco_limitacoes_final,
ifelse(Confundimento2=="Sim" |Confundimento_mediacao2=="Sim", 1, 0))
```

#Viés de seleção

```
banco_limitacoes_final$Vies_selecao3 <- with(banco_limitacoes_final,
ifelse(Vies_selecao2=="Sim" |Perda_seguimento2=="Sim"
|Dados_faltantes2=="Sim", 1, 0))
```

#Temporalidade

```
banco_limitacoes_final$Tempo3=ifelse(banco_limitacoes_final$Tempo2=="Sim", 1,
0)
```

#Transportabilidade

```
banco_limitacoes_final$Validade_externa3=ifelse(banco_limitacoes_final$Validade_
externa2=="Sim", 1, 0)
```

#Tamanho da amostra

```
banco_limitacoes_final$Tamanho_amostra3=ifelse(banco_limitacoes_final$Tamanho
_amostra2=="Sim",1, 0)
```

#Outras

```
banco_limitacoes_final$Outra3 <- with(banco_limitacoes_final, ifelse(Outra2=="Sim"
|Outra_variavel_naoobs2=="Sim", 1, 0))
```

#Tabelas descritivas

```
tapply(banco_limitacoes_final$n_limitacoes, banco_limitacoes_final$status,
summary)
tapply(banco_limitacoes_final$Avaliou, banco_limitacoes_final$status, summary)
table(banco_limitacoes_final$n_limitacoes, banco_limitacoes_final$status)
table(banco_limitacoes_final$Apenas_reconheceu, banco_limitacoes_final$status)
table(banco_limitacoes_final$Discutiu2, banco_limitacoes_final$status)
table(banco_limitacoes_final$Quantificou2, banco_limitacoes_final$status)
table(banco_limitacoes_final$Quantifi_AS, banco_limitacoes_final$status)
table(banco_limitacoes_final$Quantifi_MV, banco_limitacoes_final$status)
table(banco_limitacoes_final$Confundimento3, banco_limitacoes_final$status)
table(banco_limitacoes_final$Mensuracao_geral3, banco_limitacoes_final$status)
table(banco_limitacoes_final$Vies_selecao3, banco_limitacoes_final$status)
table(banco_limitacoes_final$Tempo3, banco_limitacoes_final$status)
table(banco_limitacoes_final$Validade_externa3, banco_limitacoes_final$status)
table(banco_limitacoes_final$Tamanho_amostra3, banco_limitacoes_final$status)
table(banco_limitacoes_final$Outra3, banco_limitacoes_final$status)
```

```
#####
```

Análises frequencia de hedges e relato das limitações

```
#####
```

```
banco_limitacoes_final$status3=as.factor(banco_limitacoes_final$status)
banco_limitacoes_final <- within(banco_limitacoes_final, status3 <- relevel(status3,
ref="Sem DAG"))
```

```
model1=lm(n_limitacoes2~ as.factor(status3), data=banco_final)
```

```
model1$coefficients
```

```
confint(model1)
```

```
model2=lm(rowSN~ as.factor(status3), data=banco_final)
```

```
summary(model2)
```

```
confint(model2)
```

```

banco_final$Quantificou3=as.factor(banco_final$Quantificou2)
fit0=logbin(reconheceu~ as.factor(status3), data=banco_final)
fit1=logbin(Discutiu2~ as.factor(status3), data=banco_final)
fit2=logbin(Quantificou3~ as.factor(status3), data=banco_final)
summary(model)
summary.logbin(fit)
summary(fit2)

```

```
#####
```

Gráficos tipo de limitações/vieses reconhecidos

```
#####
```

#Frequencia categoria “Sim”

```

banco_bias=banco_limitacoes_final[,c('status', 'Confundimento3',
'Mensuracao_geral3',
          'Vies_selecao3', 'Tempo3', 'Tamanho_amostra3',
'Validade_externa3', 'Outra3')]

```

```

a=table(banco_bias$Confundimento3, banco_bias$status)[1:2]
b=table(banco_bias$Mensuracao_geral3, banco_bias$status)[1:2]
c=table(banco_bias$Vies_selecao3, banco_bias$status)[1:2]
d=table(banco_bias$Tempo3, banco_bias$status)[1:2]
e=table(banco_bias$Tamanho_amostra3, banco_bias$status)[1:2]
f=table(banco_bias$Validade_externa3, banco_bias$status)[1:2]
status=c("Com DAG", "Sem DAG")
data_graph=(data.frame(t(rbind(status,a,a,b,c,d,e,f))))
colnames(data_graph)=c("status", "s", "Confundimento", "Viés de seleção", "Erros de
mensuração", "Temporalidade", "Tamanho da amostra",
          "Transportabilidade")
data_graph2=melt(data_graph, id="s")
data_graph3=read.csv("data_graph2.csv", header =T )

```

```
positions= c("Confundimento", "Erros de mensuração", "Viés de seleção",
"Temporalidade", "Tamanho da amostra", "Transportabilidade")
```

#Gráfico 1

```
p=ggplot(data_graph3, aes(x=variable, y =value, fill=Grupo))+
scale_y_continuous(limits = c(0,100))+
  geom_bar(stat = 'identity', position = 'dodge')+
  scale_fill_manual(values=c("blue", "lightblue4"))+
  scale_x_discrete(limits = positions)+
ggtitle("Tipos de limitações reconhecidas", subtitle="") +
  xlab("Limitações") +
  ylab("%")+
  geom_text(aes(label = value), hjust=0.5, vjust = -0.5, size=3, position =
position_dodge(width = 1), inherit.aes = TRUE)+
  theme(axis.title.y = element_text(family = "sans", color="black", size=10, hjust = 0.5,
vjust = 1),
  axis.title.x = element_text(family = "sans", color="black", size=10, hjust = 0.5,
vjust = 1),
  plot.title = element_text(family = "sans", size=14, face="bold"))
```

#Gráfico 2

```
p2=ggplot(data_graph3, aes(x=variable, y =value, fill=Grupo))+
scale_y_continuous(limits = c(0,100))+
  geom_bar(stat = 'identity', position = 'dodge')+
  scale_fill_manual(values=c("blue", "lightblue4"))+
  scale_x_discrete(limits = positions)+
ggtitle("Tipos de limitações reconhecidas", subtitle="") +
  xlab("Limitações") +
  ylab("%")+
  theme(axis.title.y = element_text(family = "sans", color="black", size=10, hjust = 0.5,
vjust = 1),
```

```
axis.title.x = element_text(family = "sans", color="black", size=10, hjust = 0.5,  
vjust = 1),
```

```
plot.title = element_text(family = "sans", size=14, face="bold"))
```

```
p2+geom_errorbar(aes(ymin =min, ymax =max),
```

```
width = 0.2,
```

```
position = position_dodge(width = 0.9))
```

```
#####
```