

Dissertação de Mestrado Profissional

**FRAMEWORK PARA ESTUDOS DE INFERÊNCIA CAUSAL EM
ESTUDOS OBSERVACIONAIS SOBRE INFECÇÕES RELACIONADAS
À ASSISTÊNCIA À SAÚDE**

ARATEUS CRYSHAM FRANCO MENESES

HOSPITAL DE CLÍNICAS DE PORTO ALEGRE
PROGRAMA DE PÓS-GRADUAÇÃO MESTRADO PROFISSIONAL EM PESQUISA CLÍNICA

**FRAMEWORK PARA ESTUDOS DE INFERÊNCIA CAUSAL EM
ESTUDOS OBSERVACIONAIS SOBRE INFECÇÕES RELACIONADAS
À ASSISTÊNCIA À SAÚDE**

Autor: Arateus Crysham Franco Meneses

Orientadora: Profa. Dra. Fernanda dos Santos de Oliveira

Coorientador: Prof. Dr. Rodrigo Pires dos Santos

Dissertação submetida como requisito parcial para a obtenção do grau de Mestre ao Programa de Pós-Graduação Mestrado Profissional em Pesquisa Clínica, do Hospital de Clínicas de Porto Alegre.

Porto Alegre

2024

CIP - Catalogação na Publicação

Meneses, Arateus Crysham Franco
FRAMEWORK PARA ESTUDOS DE INFERÊNCIA CAUSAL EM
ESTUDOS OBSERVACIONAIS SOBRE INFECÇÕES RELACIONADAS À
ASSISTÊNCIA À SAÚDE / Arateus Crysham Franco Meneses.
-- 2024.
122 f.
Orientadora: Fernanda dos Santos de Oliveira.

Coorientador: Rodrigo Pires dos Santos.

Dissertação (Mestrado Profissional) -- Universidade
Federal do Rio Grande do Sul, Hospital de Clínicas de
Porto Alegre, Programa de Pós-Graduação em Pesquisa
Clínica, Porto Alegre, BR-RS, 2024.

1. Inferência Causal. 2. Infecção Hospitalar. 3.
Causalidade. 4. Framework Causal. 5. Infecções
Associadas à Assistência à Saúde. I. Oliveira,
Fernanda dos Santos de, orient. II. Santos, Rodrigo
Pires dos, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os
dados fornecidos pelo(a) autor(a).

BANCA EXAMINADORA

Prof. Dr. Leo Sekine - Hospital de Clínicas de Porto Alegre

Profa. Dra. Vânia Naomi Hirakata - Hospital de Clínicas de Porto Alegre

Prof. Dr. José Miguel Dora - UFRGS

AGRADECIMENTOS

Aos meus pais, Hércules e Clarice, obrigado por todo o apoio e incentivo, principalmente o que definiu este período marcado pelo meu mestrado, sempre com muito amor e carinho.

A Brenda, por me apoiar e me mostrar que, apesar das dificuldades, eu era capaz de atingir meus objetivos e conquistas, obrigado.

A Profa. Fernanda dos Santos de Oliveira por ter me incentivado e orientado ao longo desta jornada, obrigado.

Ao Prof. Rodrigo Pires dos Santos e ao Prof. Tiago Andres Vaz pela mentoria não só nesta jornada mas em outras também. Vocês foram fundamentais para o meu crescimento pessoal e profissional, obrigado.

Aos meus professores do Mestrado Profissional em Pesquisa Clínica pelo aprendizado e tutoria ao longo do curso.

Aos meus colegas de curso do Mestrado Profissional em Pesquisa Clínica expresso gratidão pelo apoio com as atividades e pelo incentivo dado desde o começo deste mestrado.

LISTA DE ABREVIATURAS EM PORTUGUÊS

ANVISA - Agência Nacional de Vigilância Sanitária

CCIH - Comissão de Controle de Infecção Hospitalar

ED - Efeito Direto

EC - Ensaio Clínico

ECR - Ensaio Clínico Randomizado

EUA - Estados Unidos da América

IA - Inteligência Artificial

IRAS - Infecções Relacionadas à Assistência à Saúde

ISC - Infecção de Sítio Cirúrgico

OMS - Organização Mundial da Saúde

PLN - Processamento de Linguagem Natural

SLD - Sobrevivência Livre de Doença

LISTA DE ABREVIATURAS EM INGLÊS

CDE - Controlled Direct Effect

DAG - Directed Acyclic Graph

DE - Direct Effect

EHR - Electronic Health Records

LLM - Large Language Models

ML - Machine Learning

NHSN - National Healthcare Safety Network

OLS - Ordinary Least Square

SCM - Structural Causal Models

LISTA DE TABELAS

- Tabela 1. Taxa de mortalidade por Febre Puerperal do período de Janeiro de 1841 até Março de 1849. **61**

LISTA DE FIGURAS

■ Figura 1. Retrato desenhado de Hipócrates.	2
■ Figura 2. Retrato gravado de Aretaeus da Capadócia.	3
■ Figura 3. Gravura de James Lind.	5
■ Figura 4. Retratos de Austin Bradford Hill e Ronald A. Fisher.	6
■ Figura 5. Escada da Causalidade.	9
■ Figura 6. Exemplos de relações entre nodos em um DAG.	24
■ Figura 7. Exemplo de caminhos.	27
■ Figura 8. Exemplo de caminhos condicionais.	27
■ Figura 9. Exemplo de chain.	28
■ Figura 10. Exemplo de collider.	29
■ Figura 11. Exemplo de fork.	30
■ Figura 12. Exemplo de caminho de backdoor.	33
■ Figura 13. Exemplo hipotético de um DAG representando o tratamento de um antibiótico.	38
■ Figura 14. Exemplo aplicando as regras do do-calculus.	40
■ Figura 15. Representação do nodo de seleção.	42
■ Figura 16. Exemplo de DAGs contendo um nodo de seleção.	43
■ Figura 17. Exemplos contendo um nodo de seleção.	44
■ Figura 18. Exemplo de um estudo longitudinal para se analisar o tratamento do medicamento X em um desfecho pneumonia.	47
■ Figura 19. Exemplo para se analisar os efeitos de X em Y em um sistema linear.	53
■ Figura 20. Exemplo derivado para se analisar os efeitos de X em Y em um sistema linear.	54
■ Figura 21. Exemplo denotando os efeitos de X em Y em um sistema linear.	55
■ Figura 22. Retrato de Ignaz Semmelweis por Eugen Doby.	60
■ Figura 23. DAG contendo as potenciais hipóteses que Semmelweis fez em seu raciocínio sobre a causalidade da Febre Puerperal.	63
■ Figura 24. Intervenção através de “Lavagem de mãos” para redução dos casos de Febre Puerperal.	63
■ Figura 25. Relações entre nodos denotando a hipótese de causalidade entre “Lavagem de mãos” e o desfecho de “Mortalidade por Febre Puerperal”.	64
■ Figura 26. Tratamento do caminho de backdoor provocado pela relação entre “Contato com cadáver” e “Lavagem de mãos”.	65
■ Figura 27. Tratamento da variável “Ferida séptica”.	66
■ Figura 28. DAG demonstrando a potencial relação de causalidade da variável de “Antissepsia cirúrgica” com relação ao desfecho de “Óbito por ISC”.	69

RESUMO

Infecções hospitalares estão entre as principais causas de óbito no mundo e, em muitos casos, questionamentos sobre a origem das mesmas são de difícil constatação. O Ensaio Clínico Randomizado (ECR) sempre foi considerado a técnica padrão em estudos clínicos para elucidar relações de causalidade, porém, no contexto das Infecções Associadas à Assistência à Saúde (IRAS), este método nem sempre é adequado ou viável. Algoritmos de inferência causal surgem como uma alternativa importante para elucidar tais relações de causa e efeito, no entanto, a falta de compreensão do uso de tais métodos gera ampla heterogeneidade dos resultados, o que torna a análise dos resultados em revisões sistemáticas uma tarefa inviável. Devido a falta de padronização desses estudos, um *framework* deve servir como ferramenta fundamental para que teses de inferência causal possam emergir e, assim, proporcionar-lhes diretrizes de implementação, permitindo colaborações interdisciplinares entre estatísticos e especialistas em doenças infecciosas, levando a quebra de algumas das barreiras para a implementação de inferências causais em estudos observacionais na área das IRAS.

Palavras-chave:

Inferência causal; Infecções hospitalares; Infecções Associadas à Assistência à Saúde; *Framework* causal.

ABSTRACT

Hospital infections are among the leading causes of death worldwide, and some inquiries regarding their origin are difficult to establish. The Randomized Controlled Trial (RCT) has always been considered the standard technique in clinical studies to elucidate causality relationships; however, in the context of Healthcare-Associated Infections (HAIs), this method is not always appropriate or feasible. Causal inference algorithms emerge as an important alternative to elucidate such cause-and-effect relationships, but the lack of understanding of the use of such methods generates a wide heterogeneity of results, which makes analyzing the results in systematic reviews an unfeasible task. Due to the lack of standardization of such studies, a framework should serve as a fundamental tool for causal inference theses to emerge and, thus, provide them with implementation guidelines, allowing interdisciplinary collaborations between statisticians and infectious disease scholars, and breaking down some of the barriers to the implementation of causal inferences in observational studies in the HAI area.

Keywords:

Causal inference; Hospital infections; Healthcare-Associated Infections; Causal framework.

Sumário

1 INTRODUÇÃO	1
1.1 História da Causalidade	
1.2 Infecções Relacionadas à Assistência à Saúde - IRAS	
1.3 Tecnologia em apoio ao combate e rastreamento das IRAS	
2 JUSTIFICATIVA	18
3 OBJETIVOS	19
3.1 Objetivo Geral	
3.2 Objetivos Específicos	
4 REVISÃO DA LITERATURA	20
4.1 CONTEXTUALIZAÇÃO	
4.2 MATERIAIS E MÉTODOS	
4.2.1 Structural Causal Models (SCM)	
4.2.2 Directed Acyclic Graphs (DAG)	
4.2.3 Regras de caminho	
4.2.4 <i>do-calculus</i>	
4.2.5 Seleção de diagramas	
4.2.6 Inferência Causal e Contrafatos	
4.2.7 Inferência Causal em Sistemas Lineares	
4.3 Causalidade e IRAS	
5 RESULTADOS: Aplicação do <i>Framework</i> de Inferência Causal	59
5.1 O caso da Febre Puerperal	
5.2 Criando o DAG	
5.3 Relações entre nodos e ajustes de caminhos	
5.4 Calculando efeitos de causalidade	
5.5 Analisando contrafatos e contextos atuais	
6 DISCUSSÃO	71
7 RELATÓRIO SINTÉTICO DO PRODUTO 1 DA DISSERTAÇÃO	74
7.1 Título (Título do artigo)	
7.2 Descrição (breve, máximo 2 parágrafos)	
7.3 Aplicabilidade do produto	
7.4 Inserção social	

7.5	Previsão de publicação	
8	RELATÓRIO SINTÉTICO DO PRODUTO 2 DA DISSERTAÇÃO	75
8.1	Título	
8.2	Descrição	
8.3	Aplicabilidade do produto	
8.4	Inserção social	
8.5	Previsão de publicação	
9	ARTIGO CIENTÍFICO: “How Causal Inference Methods Could Have Aided Ignaz Semmelweis in the Puerperal Fever Case”	78
10	CONCLUSÃO	91
11	REFERÊNCIAS	95

1. INTRODUÇÃO

1.1. História da Causalidade

Dentre os precursores das grandes questões intelectuais enunciadas no nosso cotidiano existe um fundamento chamado “pensamento racional”. Precursor de um “pensamento causal”, é um processo, até então, natural realizado pelos indivíduos, moldando-lhes as decisões e orientando as ações com base nas experiências e crenças.^{1,2} Este processo, embora seja intrínseco à natureza humana, foi refletido, refinado e documentado pelos filósofos da Grécia Antiga. Desde as questões investigativas de Sócrates até as reflexões sistemáticas de Aristóteles, esses pensadores iluminaram os intrincados caminhos da razão, buscando relações e associações que levassem a um entendimento maduro e menos místico, considerando os contextos da sociedade em que viviam.^{3,4} Eles demonstraram que, através de um equilíbrio cuidadoso entre observação e introspecção, os indivíduos poderiam navegar pelas complexidades do cotidiano e chegar a uma compreensão mais clara do mundo ao seu redor.^{5,6}

Entre estes filósofos gregos que buscavam tal compreensão sobre situações cotidianas, Hipócrates teve destaque com suas contribuições marcantes na medicina (Figura 1). O filósofo era um Asclepiáde, isto é, membro de uma família que durante várias gerações praticou os cuidados em saúde, sendo considerado o “Pai da Medicina”, por defender, veementemente que as enfermidades tinham origem de “causas naturais”, contrapondo-se à ideia de origens sobrenaturais para as causas das doenças.⁷⁻⁹ Tais “causas naturais” eram elucidadas e anotadas através de observações dos enfermos ao longo do curso das doenças. A compilação dessas anotações é um documento que ficou conhecido como “Corpus Hipocrático”, onde aborda sobre a importância da observação clínica, a predição do curso das doenças e uma conduta integrada para tratar os pacientes, além das evidências e fatores associados ao surgimento das doenças e seus sintomas.¹⁰ Hipócrates enunciou a Teoria Humoral, onde sugeria que a saúde de um indivíduo era determinada pelo “equilíbrio entre sangue, catarro e bile amarela e preta”. Esta concepção influenciou

a medicina por muitos séculos, moldando as primeiras práticas terapêuticas documentadas e registradas da humanidade.^{8,11}

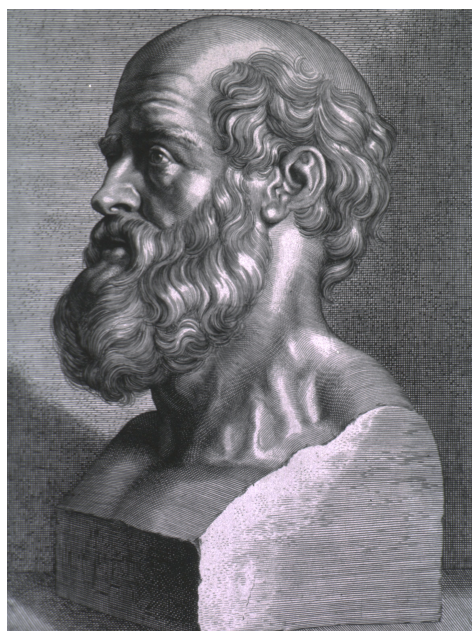


Figura 1. Retrato desenhado de Hipócrates.

Fonte: P. Pontius, 1638.

Influenciado pelas reflexões de Hipócrates, Aretaeus da Capadócia emergiu como uma das figuras médicas mais notáveis da Grécia Antiga, dando continuidade ao legado hipocrático (Figura 2).¹²⁻¹⁵ Nos seus estudos, Aretaeus ofereceu detalhadas descrições clínicas, identificou possíveis etiologias, propôs intervenções preventivas e terapêuticas, além de esboçar estratégias para combater diversas enfermidades. Classificava as doenças em categorias como "nervosas" e "mentais", bem como "agudas" e "crônicas", de acordo com sua nomenclatura específica.^{13,16-18} Entre as patologias examinadas por Aretaeus, destacam-se: pneumonia, diabetes, asma, epilepsia, tétano, carcinomas uterinos e hepáticos, histeria, cefaleia, mania, melancolia e doença celíaca.^{19,20}



Figura 2. Retrato gravado de Aretaeus da Capadócia.²¹

Fonte: Ioannes Sambucus, 1574.

A civilização grega antiga foi uma das pioneiras ao estabelecer raciocínios lógicos, causais e matemáticos. No entanto, ao compararmos seus métodos com as abordagens contemporâneas, é possível notar limitações. Fortemente baseada em contagens manuais e simples, era suscetível a imprecisões, dando margem para interpretações potencialmente equivocadas, dada a falta de rigor ou profundidade analítica.^{22–24}

Foi apenas no século 19 que os matemáticos expuseram as limitações das práticas antigas e formaram, então, métodos matemáticos e estatísticos modernos para a interpretação de números e dados. Francis Galton foi o pioneiro responsável por estabelecer relações entre os dados, determinando os conceitos de regressão e correlação. Karl Pearson aprimorou os conhecimentos de Galton e, assim, introduziu o coeficiente de correlação e o método dos momentos, sendo estas as diretrizes para os testes de hipóteses estatísticas. Em meados do século 20, Ronald A. Fisher expandiu o horizonte exponencialmente com a introdução da estimativa de máxima verossimilhança e o desenho de experimentos, estabelecendo métodos rigorosos

para fazer inferências a partir de dados. Juntos, estes estatísticos transformaram um campo de enumeração simples numa disciplina sofisticada, moldando a forma como percebemos e analisamos as complexidades do mundo na atualidade.²⁵⁻²⁷

Os métodos enunciados elucidaram questões que antes eram confundidas entre os estudiosos da época. Aquilo que se acreditava estar diretamente associado como uma relação causal de uma variável a outra, já não era mais conceituada como tal.^{24,28} As relações entre os dados se estabeleceram a fim de denotar relações ou ligação entre os fatos e, apesar de existirem relações de fato, não poderiam ser associadas como causas dos desfechos.²⁹ Estas relações, conforme fundamentado por Karl Pearson, são denominadas como “correlações”³⁰. Uma correlação corresponde a uma relação mútua entre informações ou elementos, uma dependência recíproca entre duas ou mais variáveis. Um coeficiente de correlação é uma medida numérica que mede a relação linear ou força e direção entre duas variáveis quantitativas.³¹ Denotado como a letra “r”, este coeficiente oscila entre -1 e 1, onde a representação perfeita de relação linear são os valores 1 (relacionamento linear positivo perfeito) e -1 (relacionamento linear negativo perfeito) e o coeficiente 0 denota nenhuma relação linear entre as variáveis. Um relacionamento linear perfeito não indica causa a um desfecho, visto a falta de significado que este método dá à relação entre as variáveis em seus relacionamentos umas com as outras.³²

Embora as relações estatísticas tenham sido discutidas e refinadas ao longo do século 19, é curioso notar que o método mais significativo para se associar à causa, sobretudo no contexto da área da saúde, surgiu no século 18. James Lind foi um médico escocês creditado por realizar o primeiro Ensaio Clínico (EC) da história em 1747 (Figura 3). Mesmo este estudo não tendo sido realizado seguindo todas as práticas modernas e atuais, Lind conduziu o EC de forma sistemática a fim de estudar a doença conhecida como escorbuto. Ele dividiu 12 marinheiros com os sintomas da doença em 6 grupos de dois. O primeiro grupo recebeu cidra, o segundo grupo recebeu um elixir à base de ácido sulfúrico, o terceiro grupo recebeu vinagre, o quarto grupo recebeu água do mar, o quinto grupo recebeu duas laranjas e um limão e o sexto grupo recebeu pasta apimentada e uma água de cevada. Lind observou que o grupo de número cinco teve melhora significativa dos sintomas em apenas 6 dias de consumo das frutas, enquanto os outros grupos não tiveram

melhora significativa. Com isso, James Lind associou a importância de alimentos como laranjas e limões (alimentos ricos em Vitamina C) para o tratamento de escorbuto, tendo assim melhor performance em tratar os pacientes do que os remédios da época.^{33,34}



Figura 3. Gravura de James Lind.³⁵

Fonte: I. Wright, 1783.

Tal foi o feito de James Lind que, não só pôde associar a deficiência da Vitamina C como um fator de risco para escorbuto, como também seu modelo de EC se tornou uma referência base para os estudos subsequentes na área da saúde. Porém seu modelo possuía algumas inconsistências.³³ Vieses de seleção e fatores confundidores estão entre os problemas associados ao método utilizado. Tais inexatidões levaram Pierre-Charles-Alexandre Louis em 1835 a manifestar a necessidade de uma grande quantidade de amostras (participantes de pesquisa) para uma melhor mensuração dos resultados e, assim, qualidade do estudo. Esta

contribuição ficou conhecida como “método numérico” e, a partir dela, surgiram os princípios e a necessidade de randomização em ECs.³⁶

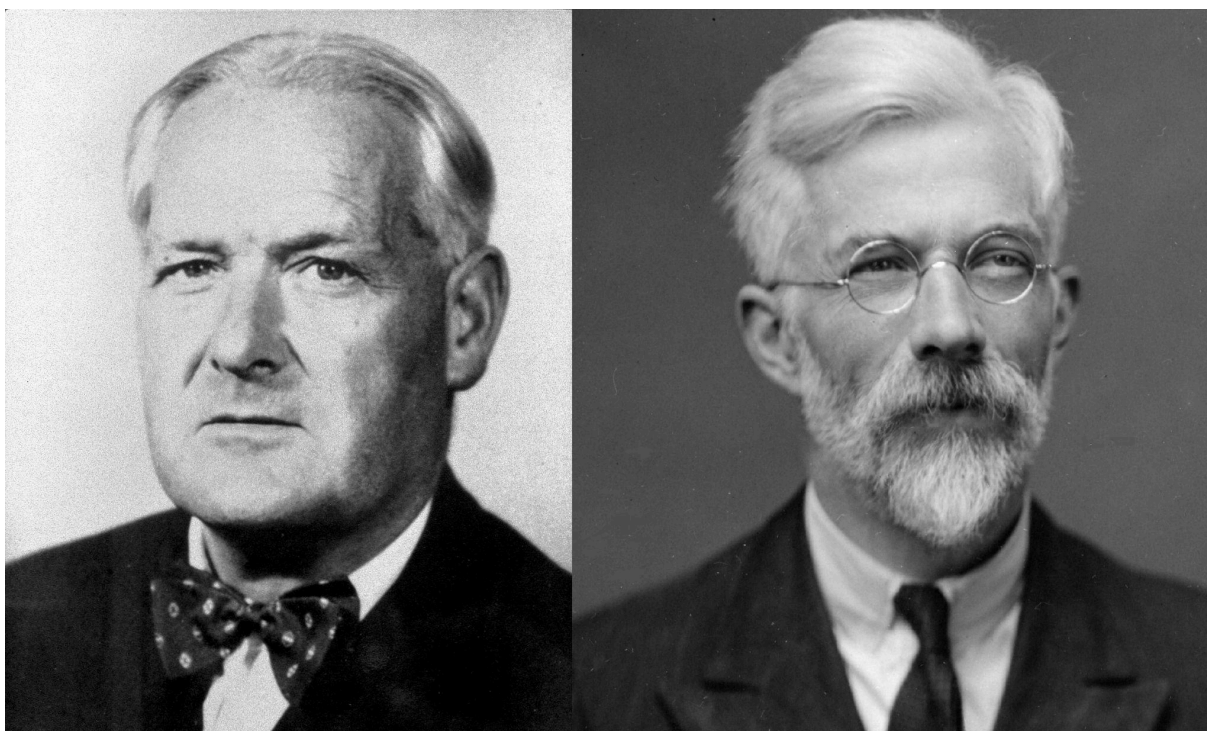


Figura 4. Retratos de Austin Bradford Hill (esquerda) e Ronald A. Fisher (direita).^{37,38}

Fontes: British Medical Journal, 1950, e Walter Stoneman, 1946.

Em meados do século 20, através das contribuições dos médicos e pesquisadores da época como Austin Bradford Hill, Ronald A. Fisher e Archie Cochrane, os ECs conduzidos adotaram uma diretriz mais robusta e consistente.³⁹ Ronald A. Fisher contribuiu significativamente do ponto de vista estatístico, fundamentando os conceitos de randomização (Figura 4). O processo de randomização busca alocar de forma aleatória as suas amostras de estudo a fim de observá-las ao longo do estudo. Seus conceitos iniciais tinham uma abordagem voltada para a agricultura, porém Austin Bradford Hill se aproveitou de tal fundamento a fim de conduzir o primeiro EC usando tal artifício (Figura 4). Com o intuito de testar a eficácia do medicamento estreptomicina no tratamento da tuberculose pulmonar, Hill selecionou pacientes de vários centros de pesquisa do Reino Unido e os distribuiu aleatoriamente em grupos. Enquanto um grupo recebeu o antibiótico estreptomicina, o outro recebeu o tratamento padrão. As avaliações dos

pacientes ao longo do processo foram registradas sistematicamente, porém os registros foram ocultados até o final da pesquisa para que vieses fossem minimizados. Após a conclusão do estudo e uma árdua revisão das anotações, a análise estatística do EC mostrou eficácia superior na melhora dos sintomas dos pacientes que fizeram o uso do medicamento estreptomicina. Desta forma, Hill não só foi capaz de oferecer avanços para o tratamento da tuberculose, como também demonstrou que o papel da randomização em ECs minimiza confundidores e vieses de estudo.^{40,41}

A contribuição de Austin Bradford Hill é amplamente reconhecida e sua metodologia para estudos clínicos foi amplamente elogiada na época.³⁹ Contudo, seus estudos subsequentes conduzidos em parceria com Richard Doll, sobre a relação entre o tabagismo e o câncer de pulmão, enfrentaram críticas significativas de alguns matemáticos e estatísticos contemporâneos. Destaca-se entre os críticos o próprio Ronald Fisher, pioneiro no uso da randomização em experimentos.^{40,42}

Hill e Doll conduziram uma série de estudos observacionais em pacientes do Reino Unido e, ao analisarem os que faziam uso de cigarro, assumiram a relação do uso do tabaco como um causador de câncer de pulmão.^{43,44} As objeções de Fisher centraram-se no fato de que, a partir da alta correlação entre esses atributos, Hill e Doll declararam que o uso do cigarro era um causador de câncer de pulmão, o que por si só não poderia ser associado como uma relação causal de fato, visto que do ponto de vista estatístico, a existência de correlação entre variáveis não é evidência suficiente de que exista causalidade entre elas.⁴⁰

Entre os argumentos de Fisher, ele enuncia fatores genéticos como potenciais confundidores e que as evidências expostas por Hill e Doll não eram suficientes para uma afirmação de causalidade. Também explicou do problema que existe em se realizar um Ensaio Clínico Randomizado (ECR) para tal análise visto que ela poderia ir contra os valores éticos e de direitos humanos.⁴² Tais críticas levantadas por Fisher foram ignoradas na época pela comunidade científica em um primeiro momento, porém seus argumentos evidenciam questões que foram arduamente debatidas na segunda metade do século 20, mostrando então como que a busca por relações de causa e efeito são aspectos de difícil coleta, análise e resolução.^{41,45}

Mesmo com as evidentes limitações apresentadas no método, o ECR foi primordialmente utilizado, desde então, como principal método para estudos na área da saúde a fim de se obter relações de causalidade em situações clínicas.⁴⁶ Estes são os primeiros métodos usados na área da saúde para obtenção de uma relação de causalidade.⁴⁴

Foi assim que técnicas causais começaram a tomar forma. A partir do entendimento de suas limitações é que métodos de inferência causal mais sofisticados foram surgindo. Por exemplo, as perguntas que não são possíveis de serem respondidas com ECRs são inseridas em um campo associado à especulação. Uma metodologia que permita estabelecer relações causais entre fatos ou contrafatos que não aconteceram, mas que sejam de interesse a serem respondidas, são fundamentais para a descoberta de novas técnicas ou consolidação dos processos já existentes.⁴⁷

Do ato de observar, conforme realizado pelos antigos gregos, até métodos que permitam antecipar os impactos de mudanças intencionais através do uso de ferramentas ou métodos (usando-se elas com um uso racional e consciente delas) e, assim, ser capaz de identificar dentre os desfechos o resultado desejado, compatível com o que foram alguns dos estudos de Bradford Hill e Ronald Fisher, se tem o que pode ser considerado os dois primeiros degraus da Escada da Causalidade (Figura 5). O domínio completo de como funcionam estes métodos e compreensão do que fazer quando estas ferramentas apresentarem falhas pode ser entendido como o terceiro degrau da Escada da Causalidade. Tal degrau requer um entendimento próximo ao que seria um raciocínio “imaginário”, onde então surgem as questões de pesquisa mais complexas e de cunho: “e se por acaso um determinado evento que não aconteceu tivesse acontecido?”.⁴⁷ Esta tese abordou sugestões para o acontecimento de estudos de tal natureza na área das Infecções Relacionadas à Assistência à Saúde (IRAS).

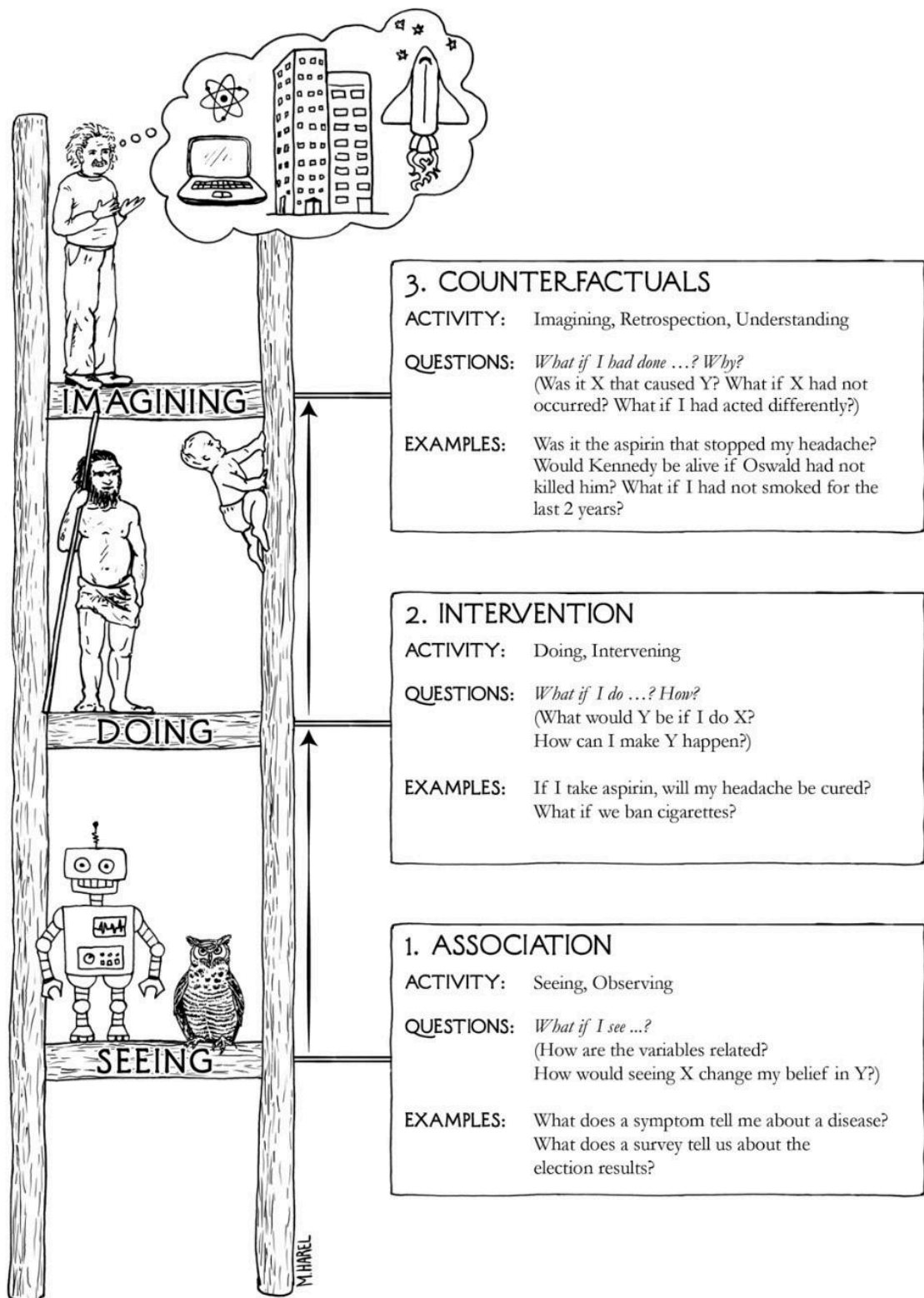


Figura 5. Escada da Causalidade.⁴⁷

Fonte: Maayan Harel, 2018.

1.2. Infecções Relacionadas à Assistência à Saúde - IRAS

As infecções hospitalares, mais recentemente denominadas IRAS, são determinantes para alta taxa de morbidade, aumento de resistência de microrganismos aos antimicrobianos, prolongamento de estadia em leito hospitalar e aumento de custos de tratamento para pacientes internados em um hospital.⁴⁸ Por definição, são consideradas IRAS as infecções que acontecem após as primeiras 48 horas da admissão hospitalar ou após a alta, quando puder estar relacionada com a internação ou procedimentos hospitalares prévios.⁴⁹

A fim de tentar buscar melhor prevenção e manejo para pacientes com diagnósticos de infecções hospitalares, alguns hospitais têm instituído grupos que fazem o estudo e controle das IRAS.⁴⁸ Estima-se que um terço das IRAS podem ser prevenidas quando se institui um programa de controle de infecção local.⁵⁰

Nos Estados Unidos da América (EUA), onde existem sistemas de vigilância de controle de infecções hospitalares bem estabelecidos, estima-se que 3% dos pacientes hospitalizados em 2015 tenham tido uma ou mais IRAS. Neste ano de 2015 foi estimado uma média de 687.000 infecções e, aproximadamente, 72.000 destas pessoas infectadas evoluíram para um desfecho de óbito.⁵¹ Mesmo sabendo da existência dos diversos fatores associados à mortalidade, como por exemplo pressão alta, obesidade, uso de tabaco, diabetes, doenças sexualmente transmissíveis, cânceres etc;⁵² as infecções hospitalares nos EUA se situam entre as 10 principais causas de morte no país.⁵³

Em países considerados subdesenvolvidos e emergentes as taxas de IRAS são ainda maiores, quando comparadas àquelas dos países desenvolvidos. No Brasil, desde 1999 a Agência Nacional de Vigilância Sanitária (Anvisa) vem exercendo o papel de coordenar e apoiar tecnicamente as Coordenações Distritais/Estaduais e Municipais no controle e redução das IRAS.⁵⁴ Estes programas propostos levaram o país a um aumento do percentual de conformidade de adesão às boas práticas de prevenção de IRAS, seguindo as diretrizes desenvolvidas e recomendadas em conjunto com a Organização Mundial da Saúde (OMS), levando a uma redução dos índices de infecções hospitalares.^{55,56}

A definição das taxas de infecção de determinada instituição é realizada através da revisão de casos em busca de critérios para o diagnóstico epidemiológico das IRAS. Essa atividade é essencial e uma das principais demandas das equipes de profissionais de saúde que compõem as Comissões de Controle de Infecção. Recomenda-se que esta atividade seja realizada através de uma busca ativa por critérios epidemiológicos definidos pelo CDC/Anvisa.⁴⁹ A partir da análise e identificação de infecções realizada pelos controladores de infecção, que é feita de forma padronizada e estruturada mensalmente, através desta vigilância epidemiológica, os centros e instituições de saúde podem tomar decisões mais objetivas e estabelecer prioridades de ação mais assertivas para prevenção e até eliminação das IRAS.⁵⁷

Historicamente, nos hospitais brasileiros, a principal metodologia utilizada para a vigilância epidemiológica tem sido a realizada de forma manual. Este método se baseia em uma revisão de diversos dados do paciente, dentre eles: revisão de notas de evolução em prontuários na busca de sinais e sintomas de infecção, dados microbiológicos, uso de antibióticos, avaliação de exames de laboratório e imagem⁴⁹. Nesta revisão, busca-se identificar uma série de critérios de infecção, que são representados por conjuntos de regras e normativas definidos por entidades internacionais para auxiliar profissionais no controle de infecções hospitalares.^{57,58}

Apesar desta metodologia seguir uma série de normativas, regras e critérios específicos da Anvisa, derivados de programas do Centro de Controle de Doenças americano, OMS e *National Healthcare Safety Network* (NHSN), sabe-se da dificuldade que é realizar uma operação de monitoramento de infecções hospitalares em todos os pacientes internados no hospital.⁴⁸ O processo de busca por infecções hospitalares em hospitais pode depender da disponibilidade de profissionais qualificados para realizar o processo; da qualidade e quantidade de medidas preventivas adotadas e adequadamente implementadas; e as potenciais nuances entre os fatores metodológicos podem variar conforme aspectos culturais na busca por infecções de cada organização ou unidade hospitalar.^{59,60}

Tal processo, mesmo quando é realizada a busca por exames culturais positivos, demanda uma grande quantidade de tempo do grupo de pessoas que

realiza o controle de infecção.⁶¹ Estima-se um gasto semanal de 18 horas por semana em média de trabalho realizado por um profissional do controle de infecção para cada 100 leitos hospitalares na busca ativa de infecções hospitalares.⁶² Tempo esse que o profissional poderia despender em atividades, por exemplo, de consultorias, visitas às unidades e educação.⁶¹

Apesar deste contexto padronizado para o diagnóstico ou classificação de um evento infeccioso como IRAS, os fatores que levam a infecções são complexos. Questões relacionadas ao paciente, como comorbidades, imunossupressão, tabagismo; fatores associados à instituição como a quantidade e qualidade de medidas preventivas adotadas ou fiscalizadas, perfil de resistência bacteriana, contribuem diretamente com as taxas de infecção em unidades clínicas e hospitais brasileiros. Ainda, nuances operacionais de uma instituição nas relações de pacientes/profissionais da saúde, tais como higiene de mãos, também estão associadas à uma mudança nas taxas de infecção hospitalar.^{60,63,64} Mesmo estes fatores sendo conhecidos em estudos sobre infecções hospitalares, existe certa incerteza em estabelecer ou comprovar de fato quais destes estão diretamente associados como fatores causais para o acontecimento das IRAS.^{59,65-69}

1.3. Tecnologia em apoio ao combate e rastreo das IRAS

As detecções das IRAS são beneficiadas por inovações tecnológicas capazes de consultar e consumir dados eletrônicos em saúde.⁷⁰ Softwares que analisam dados ou fazem o uso de Inteligência Artificial (IA) demonstram grande potencial para automatizar e otimizar o processo de vigilância às IRAS, reduzindo drasticamente o tempo necessário para identificar infecções.⁷¹⁻⁷³ Através de algoritmos avançados, sistemas baseados em modelos de *Machine Learning* (ML) podem analisar rapidamente grandes conjuntos de dados clínicos, identificar padrões e, desta forma, adquirirem a habilidade de executar tarefas similares às aquelas realizadas em processos humanos.^{60,74,75}

Tão relevante quanto o processo de detecção de IRAS, as novas ferramentas baseadas em dados computacionais podem auxiliar na identificação de fatores de

risco, tendências de resistência a antimicrobianos e potenciais causas para o acontecimento das IRAS. Desta forma, tais métodos podem não só aumentar a eficiência do processo de vigilância às IRAS, mas também contribuir para uma melhoria contínua nas práticas de controle de infecção. Desempenham um papel importante na melhoria dos cuidados e segurança dos pacientes, identificando padrões e prevendo possíveis desfechos.⁷⁶

A importância destes sistemas informatizados é notória. A cultura de trabalho em um ambiente hospitalar se beneficia fortemente com a introdução de IA, *Big Data* e métodos de Inferência Causal.^{72,73} Em aspecto análogo, os dados são coleções digitais de eventos e observações e, no contexto da saúde, são essenciais para a prestação de cuidados com os pacientes, realização de estudos observacionais e ML, por exemplo. Estas evidências ou fatos eram registrados e anotados em prontuários de papel. Nos tempos atuais, quando registradas em sistemas computacionais, são denominados como *Electronic Health Records* (EHR).^{77,78}

As contribuições que envolvem inovação e o uso de técnicas de Inteligência Artificial transitam em vários setores. No âmbito da área da saúde, as oportunidades transitam de processamento de imagens de Raio X, como por exemplo classificação de pneumonia usando-se exames de Raio X de tórax,⁷⁹ a auxílio robótico em procedimentos cirúrgicos, como por exemplo, o procedimento para transplante de rins.⁸⁰

Diversos estudos envolvendo processamento de texto, usando-se técnicas de Processamento de Linguagem Natural (PLN) ou *Word Embeddings* surgem com grande impacto positivo na área da saúde. Modelos capazes de identificar expressões regulares sofisticadas em texto apoiam entidades hospitalares a identificar pacientes com alto risco de estarem com infecções hospitalares;⁸¹ da mesma forma que algoritmos a base de Redes Neurais (*Deep Learning*) surgem como potenciais classificadores de entidades em texto, permitindo, assim, que modelos de IA adquiram uma habilidade de analisar e compreender contextos em que uma palavra se encontra em uma frase presente, por exemplo, em um prontuário eletrônico de um hospital.⁸²

Nas circunstâncias atuais em que se encontram esses adventos tecnológicos baseados em processamento de texto também vale ressaltar o espaço crescente que os algoritmos de *Large Language Models* (LLM) vêm adquirindo. Com contribuições para novas descobertas científicas em todos os espectros da ciência moderna, estes modelos apresentam um novo capítulo na área da IA. Entre as possibilidades de atuação destes modelos estão a capacidade de fazerem análises de texto sofisticadas para fins de auxílio de diagnóstico e de tratamento de doenças ou até mesmo gerar notas sintéticas de evoluções hospitalares para treinamento de Redes Neurais.⁸³⁻⁸⁵ Contudo, existem acadêmicos que fazem críticas recorrentes a estes modelos. O cerne de tais desaprovações focam no fato de que a geração de informações realizadas pelas LLMs podem ser factualmente incorretas ou enganosas, conduzindo a riscos a longo prazo, incoerências éticas e, assim, levando a consequências graves.⁸⁶

No contexto da infectologia hospitalar, existem produtos no mercado que fazem uso de IA para apoiar clínicos e profissionais que integram as Comissões de Controle de Infecção Hospitalar (CCIH).^{73,87-90} Dentre alguns destes produtos e inovações existem o reconhecedor de entidades nomeadas e o “LLM da saúde” da John Snow Labs, que fazem uso de técnicas que envolvem PLN e LLM;^{91,92} Sistemas automatizados e semiautomatizados agregam em qualidade para a classificação e no tratamento de pacientes hospitalizados com IRAS na Europa e Brasil.⁷²

Estas contribuições se mostram como fortes reforços para os profissionais da saúde, agregando em na efetividade da vigilância.^{73,87,88,93} Buscando uma melhora na qualidade do tratamento que é ofertado aos pacientes, podem auxiliar na redução de esforços na busca ativa de infecções, reduzir potenciais erros na identificação de casos de infecção, aumento da efetividade da vigilância e, conseqüentemente, redução das taxas de mortalidade e custos associados à internação de um paciente com infecção hospitalar.^{72,94-98}

Manuais de vigilância são exigentes para quem os segue, possuem pouca padronização em como segui-los, tornando a atividade de controle de infecções uma tarefa custosa em questão de tempo em que se leva a realizar tal ofício. Algoritmos

de IA podem dar suporte no desenvolvimento de algoritmos que permitam o entendimento dos fatores de riscos associados às IRAS; aumentando a estratificação de risco, identificação dos caminhos de transmissão e até mesmo a detecção periódica ou em tempo real das infecções.^{72,73}

Apesar de existir um crescente número de pesquisas utilizando ML no campo da saúde, estes métodos têm limitações quanto à exposição das causas dos resultados. Em diversas situações, os modelos de ML e *Deep Learning* funcionam como "caixas pretas", sem oferecer um entendimento claro e transparente dos processos que conduzem a tais desfechos.⁹⁹

Historicamente, para poder se estabelecer uma relação causal a desfechos, conforme mencionado anteriormente o ECR tem sido o principal método usado para este fim na área da saúde, pois ele minimiza vieses e fatores de confusão de estudo.^{46,100} Porém, por conta de relações éticas, potenciais resultados perigosos, longos períodos para realização do estudo, condições crônicas entre os pacientes e altas taxas de evasão de participantes, este método pode acabar não sendo o modelo mais apropriado para se estabelecer as relações causais entre desfechos.^{101,102}

O *framework* de inferência causal fundamentado por Judea Pearl surge como um grande contribuinte para estudos de causalidade e de ML na área da saúde. O maior objetivo de diversos estudos clínicos é estimar a causalidade dos eventos que levam aos desfechos.¹⁰² Permitindo relacionar e articular as relações causais em estudos observacionais através de fórmulas matemáticas, pode-se ter um melhor entendimento dos fatos e, assim, agregar nas sistemáticas existentes uma habilidade de expressar informações relacionadas que sejam de desejo serem encontradas ou identificadas.⁴⁷

A busca por causas a desfechos é fruto de estudos pelos seres humanos por mais de três séculos.¹⁰³ A área da inferência causal pode ser definida como o estudo e implementação de técnicas que permitem aos pesquisadores fazerem conclusões causais a partir de dados científicos, sem ter a necessidade de depender apenas de associações ou distribuições.^{104,105} Diagramas Causais, também conhecidos como *Directed Acyclic Graphs* (DAGs), surgem como ferramentas poderosas para

representar e analisar as relações causais entre as variáveis.¹⁰⁶ Porém, o DAG por si só não é capaz de comprovar a causalidade entre os eventos. Desta forma, o uso do método do *do-calculus* e o uso de contrafactuais se torna fundamental para embasar e elucidar as relações causais.⁴⁷

O *do-calculus* é a representação matemática dos diagramas causais. Contendo uma série de regras e algoritmos matemáticos, este componente é o fator determinante para permitir que pesquisadores possam estabelecer as probabilidades condicionais entre os efeitos causais.^{47,107}

Desta forma, a psicologia enuncia o pensamento contrafactual como um conceito que evolui como a tendência de criar alternativas possíveis a eventos que já tenham ocorrido e que possam responder a questionamentos. A utilização de um modelo causal em conjunto com um método de inferência transcende os métodos tradicionais da estatística. Esse enfoque vai além das abordagens convencionais, incorporando uma análise com fins de identificar potenciais relações causais e extrair inferências significativas. Técnicas contrafactuais se estabelecem por explicar as causas, eventos, atribuições e “políticas” probabilísticas relacionadas que levam a diversos desfechos determinados relacionados a questões ou situações que buscam ser entendidas.⁶

Aplicando-se as regras e algoritmos do *do-calculus* dentro de um *framework* em conjunto com suas representações em DAGs, pesquisadores podem obter modelos causais capazes de mensurar relações causais entre as variáveis. Desta forma, se torna possível responder a perguntas causais e contrafactuais que, hoje, são fundamentais para entendermos os dados e como os eventos se sucedem, o que pode ser considerado como informações de imenso valor.¹⁰⁸

Comumente, estudos observacionais focam em estabelecer associações entre variáveis de interesse e seus desfechos para conseguir determinar as relações causais em um modelo de inferência causal. A falta de entendimento de pesquisadores com relação ao campo da inferência causal tem se demonstrado uma situação problemática. Erros sistemáticos, vieses secundários e confundidores que não são identificados ou manejados utilizando-se métodos estatísticos adequados têm sido recorrentes nos estudos de inferência causal por falta de padronização

entre os métodos, aumento das cargas de dados (*Big Data*), compreensão precária dos vieses e confundidores e, principalmente, má aplicação dos novos métodos para inferir as causas.¹⁰⁹

Hufstedler (2022) em sua revisão sistemática de estudos de inferência causal na área das doenças infecciosas apontou que a falta da implementação de métodos causais modernos em estudos observacionais relacionados a doenças infecciosas tem apresentado diversas inconsistências. A heterogeneidade na condução e implementação dos estudos, devido à falta de padronização, impediu comparações efetivas entre as pesquisas.¹¹⁰

Devido a essa questão, Hufstedler (2022) sugere a criação de manuais e guias que divulguem, de forma transparente, métodos rigorosos para maior padronização dos estudos. Tais diretrizes devem orientar pesquisadores a implementar processos que levem a uma maior harmonização entre os resultados das teses de inferência causal, permitindo que o leitor possa, desta forma, compreender adequadamente como as análises foram conduzidas.¹¹⁰

2. JUSTIFICATIVA

As IRAS continuam a ser causa de significativa de morbidade em pacientes hospitalizados demandando a identificação de suas causas subjacentes. Embora ensaios clínicos sejam uma abordagem tradicional para essa investigação, torna-se financeiramente inviável ou eticamente impraticável conduzi-los em pacientes hospitalizados por causas primárias.¹¹¹

Estudos de inferência causal, por outro lado, oferecem uma alternativa ética e acessível para se estudar relações causais. Ao serem empregados de maneira apropriada, esses estudos têm o potencial de serem instrumentos essenciais na detecção das causas das infecções hospitalares quando ECRs não são viáveis.⁴⁷

Contudo, diante da atual falta de compreensão entre pesquisadores sobre como elaborar e conduzir estudos de inferência causal, urge uma abordagem científica mais aprofundada nesse método.¹¹⁰ Propõe-se, assim, a criação de um *framework* abrangente contendo técnicas e um Diagrama Causal fundamentado na literatura. Este *framework* servirá como um guia para a realização de projetos sobre métodos de inferência causal e contrafactual na área das IRAS, não apenas viabilizando estudos nesse domínio, mas também estabelecendo um ambiente propício para essas pesquisas.

O presente estudo buscou estabelecer critérios para a implementação dessas técnicas. Dessa forma, as pesquisas derivadas deste *framework* poderão validar as relações causais propostas pelo Diagrama Causal e, adicionalmente, criarem novas relações causais para descobrir ou mensurar outros fatores associados. A metodologia proposta visa não apenas desenvolver modelos matemáticos para estudos de inferência causal na área das IRAS capazes de simular cenários interventivos em dados observacionais, mas também apresentar um Diagrama Causal contendo as relações causais extraídas da literatura, oferecendo suporte aos pesquisadores na condução de seus estudos de causalidade.

3. OBJETIVOS

3.1 Objetivo Principal

Desenvolver e disponibilizar ao meio acadêmico um *framework* para estudos de inferência causal na área das IRAS capaz de simular cenários interventivos. Este *framework*, contendo técnicas, sugestões de boas práticas e um Diagrama Causal, visa preencher uma lacuna no conhecimento acadêmico, proporcionando uma ferramenta valiosa para pesquisadores engajados em estudos de inferência causal nas IRAS.

3.2 Objetivos Específicos

- Identificar e estabelecer as técnicas e boas práticas recomendadas para a realização de estudos de inferência causal na área das IRAS, com base na revisão da literatura.
- Desenvolver um DAG associado a um caso de IRAS, incorporando as principais relações de causalidade e interações complexas com base nos critérios identificados na revisão da literatura. O diagrama deve servir como uma ferramenta visual para auxiliar na compreensão base para a condução de estudos de inferência causal.
- Desenvolvimento de um artigo científico elaborando o caso de “Ignaz Semmelweis e lavagem de mãos no combate de Febre Puerperal” utilizando-se de técnicas de inferência causal modernas, demonstrando como tais técnicas poderiam ter auxiliado as hipóteses de Semmelweis na sua época e como, em tempos modernos, tais técnicas podem auxiliar pesquisadores a elucidar relações de causa e efeito.

4. REVISÃO DA LITERATURA

4.1. CONTEXTUALIZAÇÃO

Muitos dos questionamentos dos que estudam epidemiologia e de pesquisadores que realizam estudos clínicos consistem em determinar como a saúde de alguém varia conforme o passar do tempo.¹¹² O processo de compreensão do histórico de uma doença na linha de tempo da humanidade e a capacidade de prever-se um prognóstico é objeto de grande relevância para clínicos e pacientes. Apesar de tal pesquisa ter tamanha importância na área da saúde, este estudo e análise é de grande complexidade quando realizado em dados observacionais.^{113–115}

Os ECRs são frequentemente vistos como o método mais confiável na medicina para estabelecer relações causais. Quando um ECR é somado a dados observacionais, gera-se um benefício significativo ao estudo, como melhora da validade externa, a possibilidade de integrar rapidamente uma grande quantidade de dados e a capacidade de explorar um número maior de hipóteses de estudo que podem ser testadas.^{116–118} Ainda sim, existem situações onde um ECR não pode ser praticado, seja por ser inapropriado ou por motivos éticos, tornando estudos observacionais a única abordagem possível para delineamentos clínicos.¹¹⁰

Estudos observacionais longitudinais de doenças infecciosas têm sido beneficiados pelo uso de métodos de inferência causal.¹¹⁹ A possibilidade de se acompanhar os fatores que precedem o surgimento de uma doença através de uma série de eventos, condições ou características registradas em formato de dados eletrônicos, por exemplo, abre um leque de opções e métodos que podem ser testados a fim de se buscar relações causais.¹²⁰

O uso de métodos de inferência causal, embora tenham se mostrado eficazes em estudos epidemiológicos e de doenças infecciosas, comumente possuem grandes incertezas sobre a qualidade dos resultados que são obtidos a partir deles. A falta de padronização no uso dos métodos utilizados implica em possíveis falhas na utilização dessas técnicas, gerando imprecisão nos resultados. Desta forma, a

compreensão dos métodos e técnicas é fundamental para a realização de estudos causais modernos em dados observacionais.^{110,119}

4.2. Materiais e Métodos

4.2.1. Structural Causal Models (SCM)

Uma definição para relações de causalidade pode ser dada por: uma variável X é uma causa direta de uma variável Y se X aparecer na função que atribui o valor de Y . A variável X é uma causa da variável Y se for causa direta de Y ou de qualquer causa de Y .¹²¹ Quando se realizam estudos de causalidade, comumente existem variáveis e desfechos que são prioritários ou de maior interesse a serem estudados. Descrevê-las de forma sucinta e ter a capacidade de explicar como as características relevantes interagem entre si é fundamental para o entendimento e validade das relações causais. Ao se aplicar uma organização estrutural a um conjunto variáveis relevantes, se cria um entendimento quantitativo de como os fatores se relacionam no escopo de estudo através de uma função. Podendo ser associadas a um grafo para se obter ainda um entendimento qualitativo, se tem o que pode ser denominado como um *Structural Causal Models (SCM)*.^{122,123}

Um SCM deve permitir aos pesquisadores a capacidade de representar as suas suposições causais em um formato de fácil compreensão. A habilidade de simbolizar como as relações causais se estabelecem e, como se pode tratar contrafactuais considerando o seu escopo de estudo, é uma das grandes qualidades de se ter um SCM.⁴⁷

As suposições causais criadas por pesquisadores também podem ser denominadas como “Teoria Causal”. O SCM pode não só ser um método para representar como as variáveis de interesse se comportam entre si, mas também permite assumir características externas (ou naturais) que sejam de complexa análise ou mensuração, como fatores perturbadores ao objeto de estudo. A importância de se considerar estas variáveis externas em um modelo causal é

tamanha pois elas podem alterar a forma como os vínculos do modelo se estabelecem, alterando as relações de causa e efeito entre os nodos, mesmo que sejam mutuamente independentes.¹²²

Com isso, um SCM é um grupo de variáveis U e V e uma função f que atribui um valor a cada variável V baseado nos valores das outras variáveis do modelo. As variáveis U em um grafo são denominadas como nodos raiz. Não podendo ser descendentes de quaisquer outras variáveis, são designadas como “externas” ao modelo. Estas variáveis U são chamadas de variáveis exógenas. As variáveis V que são descendentes de pelo menos uma variável exógena, são denominadas como variáveis endógenas.¹²¹

Um aspecto importante que deve ser considerado ao se criar um gráfico causal é considerar fatores adicionais denominados como “fatores omitidos”. Estas variáveis exógenas são fatores que não foram medidos pelo fato de não ser possível serem mensurados ou porque são considerados irrelevantes para o modelo. Porém, isto não quer dizer que eles não afetem diretamente as variáveis endógenas do modelo. Algumas técnicas aplicadas aos gráficos causais auxiliam a detectar quando fatores omitidos são relevantes ou não para o entendimento dos SCMs (conteúdo que será visto nas etapas seguintes).⁶

Como referido, uma forma de representar um SCM é através de um modelo de gráfico causal. Estes gráficos, denominados como *Directed Acyclic Graphs* (DAG), permitem compreender as relações entre as variáveis e dar uma representação de causalidade entre elas.^{6,124,125}

4.2.2. Causal graph analysis - DAG

A representação visual qualitativa das relações causais que descreve as relações funcionais entre as variáveis de interesse é chamada de *Directed Acyclic Graphs* (DAG). Além de permitir uma visualização das relações entre as variáveis e covariáveis, um DAG permite analisar o comportamento de tratamentos e os potenciais desfechos clínicos associados. Este método gráfico também permite ao

pesquisador identificar variáveis confundidoras, diminuir vieses de estudo e, assim, gerar um modelo estatisticamente apropriado.^{109,126-128}

Cada nó de um DAG representa uma variável. A questão de pesquisa é denotada pela relação causal de interesse. Para que uma variável tenha uma associação de relação causal, esta variável, denominada como exposição ou pai, precisa necessariamente ter uma associação causadora de outra variável, neste caso conhecida como resultado, desfecho, descendente ou filho. A variável de exposição também pode ter uma relação de causalidade caso ambas compartilhem uma causa de origem em comum.¹²⁹ Uma seta (ou arco) unidirecional indica a relação de efeito que a exposição exerce sobre o resultado, não podendo existir retorno de volta desta variável para si mesma (acíclica) e todos os arcos devem ser direcionados para uma única direção. Nesta representação gráfica, a variável pode ser contínua, ordinal, categórica ou binária.¹²⁴ Quando uma variável X (exposição) interfere ou transmite diretamente algum efeito a uma variável Y (desfecho), temos a mais simples de uma relação causal (Figura 6A).^{122,123,130}

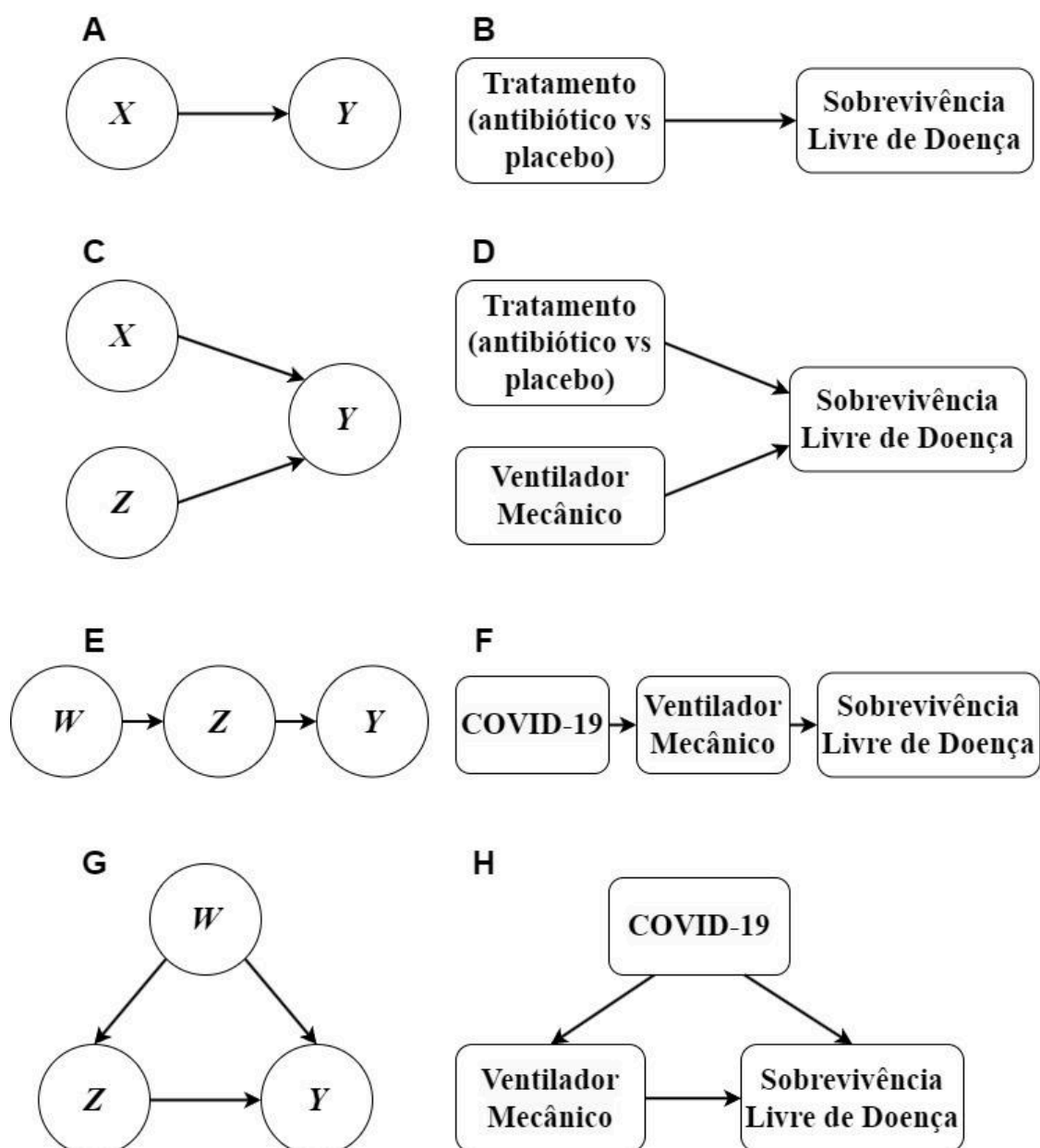


Figura 6. Exemplos de relações entre nodos em um DAG.

Fonte: O autor, 2024.

Em um exemplo na área da infectologia em um paciente com Pneumonia, como demonstrado nas Figuras 6A e 6B, o uso do antibiótico X ($X = 1$) ou placebo ($X = 0$) age de forma direta ao tempo de recorrência da doença ou sobrevida do paciente. Esta relação de Sobrevivência Livre de Doença (SLD) é definida como $Y = \text{Tempo de SLD}$.^{107,131}

Em uma situação onde outra variável Z , como o uso de ventilador mecânico por exemplo, seja adicionada de forma a afetar diretamente a variável Y (Tempo de SLD), ambas as variáveis X e Z afetarão o desfecho Y (Figura 6C, D). Neste caso, a variável Y será denotada como uma variável *colisora* (“*collider*”), pois neste caso, ela atua de forma a bloquear o caminho das variáveis X e Z .¹³²

Caso se tenha a existência de uma nova variável W , como por exemplo a presença de infecção por SARS-CoV-2 (COVID-19) neste caso de um paciente com Pneumonia, e este novo fator seja um dos motivos pelos quais foi feito o uso de ventilador mecânico (variável Z),¹³³ a relação entre a exposição W e o desfecho Y será mediado pela variável *Mediadora* (“*mediator*”) Z (Figura 6E, F).¹²³ Os nodos filhos e seguintes gerados a partir da exposição W são chamados de *descendentes* (“*descendant*”), neste caso da Figura 6E, F são as variáveis Z e Y , e implica em transmitir a relação de causalidade. Dependendo do contexto, estas associações de causalidade podem ser diretas ou indiretas. Em contraponto, variáveis afetadas diretamente ou indiretamente por outra variável são denominadas como *ancestrais* (“*ancestors*”).^{131,134,135} Neste exemplo das Figuras 6E e 6F os ancestrais de Y são Z e W .

Por fim, caso se estabeleça uma ligação entre as exposições de COVID-19 W , uso de ventilador mecânico Z e, ainda, se atribuir a variável W uma relação ao desfecho de pneumonia Y (neste caso, uma “pneumonia associada ao uso de ventilador mecânico”), a variável W funcionará como uma variável *confundidora* (“*confounding*”).^{106,123} Desta forma, a variável W influenciará tanto no desfecho de Tempo de SLD Y quanto no uso de ventilador mecânico Z , como denotado nas Figuras 6G e 6H.

Note que é importante perceber que, para gerar o DAG, se deve ter o entendimento de que um DAG não só representa as relações causais de um SCM, mas também conta a interpretação de uma “história causal”. A interpretação de quem gera o gráfico causal afeta diretamente o modelo causal final.¹³⁶

Caso mais variáveis forem adicionadas ao DAG, mais complexo se torna interpretar o diagrama causal. Caso o gráfico final seja muito complexo, uma alternativa pode ser em reduzi-lo a uma descrição causal mais simples e direta. Essa

descrição deve permitir ao pesquisador atingir seus objetivos de estudo e expressar as relações causais de desejo sem ser prejudicado por um DAG excessivamente complexo.^{137,138}

Os exemplos demonstrados na Figura 6 são apenas para fins didáticos. Em uma situação real, a importância de se contemplar uma relação adequada de variáveis é fundamental para se atingir o diagrama causal desejado.¹⁰⁷ Caso o modelo causal demonstre incoerências, o mesmo deve ser refinado a fim de se identificar quais as variáveis que devem ou não constar no DAG. A presença de determinados fatores pode variar conforme cada população de estudo e objetivo da tese em si.¹³⁹⁻¹⁴¹

4.2.3. Regras de caminhos

Quando já se tem uma seleção de quais variáveis irão compor um DAG, saber como que elas serão distribuídas e como que as arestas se estabelecerão no gráfico é fundamental para representar apropriadamente o modelo causal. Lembrando de que a função do modelo causal é identificar confundidores e diminuir vieses, pode-se atingir um modelo que seja estatisticamente apropriado e que permitirá ao pesquisador responder algumas de suas perguntas de pesquisa.¹²⁸

Relações causais em um DAG são representadas por caminhos. Um caminho é uma sequência adjacente entre variáveis conectadas por arestas direcionais que não estejam bloqueadas.¹²⁷ Sem existir caminhos fechados ou cíclicos, estas conexões demonstram como as relações entre os fatores estão condicionadas, permitindo a quem observa o gráfico perceber se seu modelo causal está correto ou inadequado.¹²¹

Se em um caminho de um DAG, duas ou mais arestas distintas apontarem para um mesmo nodo, este caminho está *bloqueado*. Este nodo é denominado como um *collider*. Caso esta característica não ocorra, o caminho está *desbloqueado*.¹⁴¹ O bloqueio em um nodo atua de forma a “romper” a transmissão de associação no percorrer do caminho. A Figura 7 exemplifica nodos que podem bloquear ou

desbloquear caminhos. É importante ressaltar que algumas outras características e técnicas podem bloquear ou desbloquear caminhos de um DAG, viabilizando associações de causalidade.¹³⁴

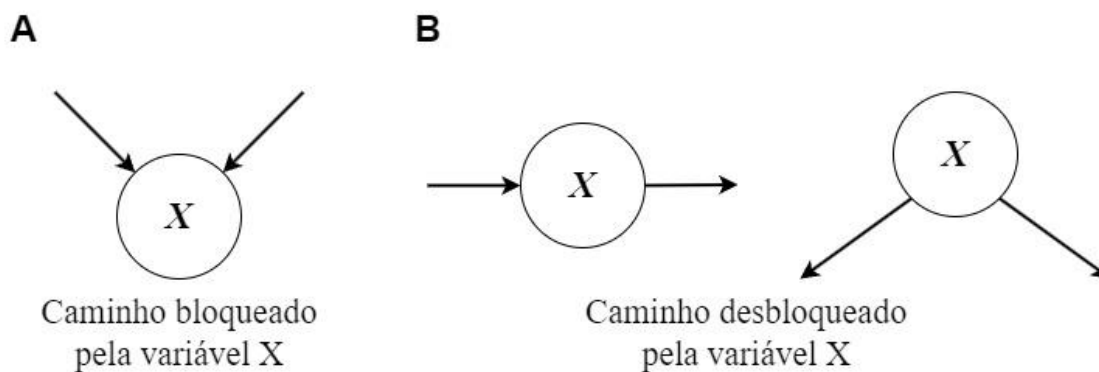


Figura 7. Exemplo de caminhos.

Fonte: O autor, 2024.

Em contraponto a Figura 8 mostra que, quando uma variável é *condicional*, ela atua de forma a inverter a relação de bloqueio ou desbloqueio da variável em um nodo de um DAG. A variável X na Figura 7A atua como uma variável *colisora* (“*collider*”), bloqueando o caminho do DAG. Quando condicional atua de forma a desbloquear o caminho (Figura 8A). Assim, na Figura 7B, a variável X que antes era desbloqueada, quando condicional, atua de tal forma a bloquear o caminho (Figura 8B).^{47,132}

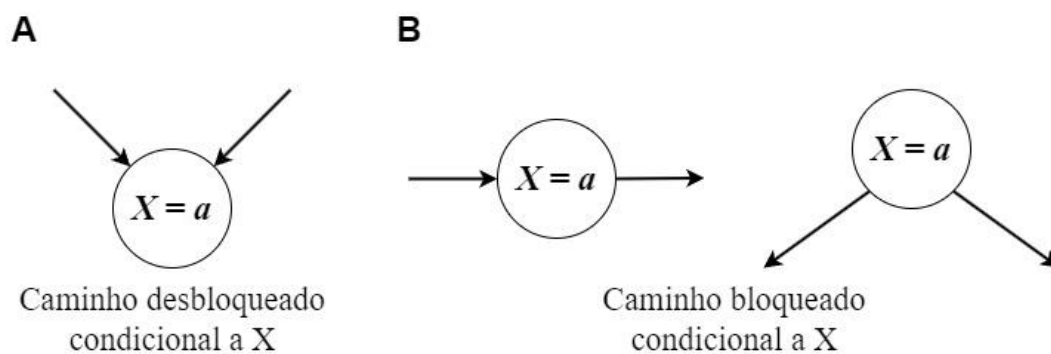


Figura 8. Exemplo de caminhos condicionais.

Fonte: O autor, 2024.

Para constatar o comportamento das variáveis representadas em nodos conectados em um gráfico causal, é necessária a compreensão de dependência ou independência entre características. Uma relação de dependência entre variáveis se estabelece quando uma variável X é capaz de influenciar ou alterar o valor da variável Y , de acordo com que X varia.¹⁴² A Figura 9 demonstra uma relação de *chain* um gráfico causal que contém as prováveis relações de dependência:

1. Z e Y são provavelmente dependentes;
2. Y e X são provavelmente dependentes;
3. e, logo, Z e X também são provavelmente dependentes.

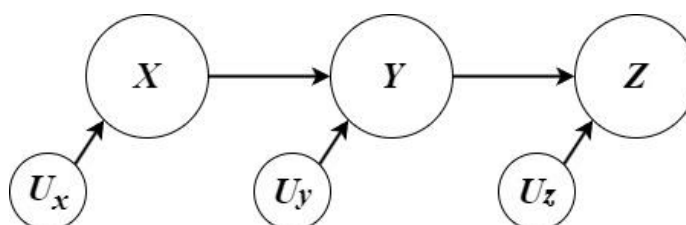


Figura 9. Exemplo de *chain*.

Fonte: O autor, 2024.

Neste gráfico causal representado pela Figura 9, as variáveis exógenas U (U_x , U_y , U_z) representam fatores desconhecidos ou aleatórios independentes entre si. Sendo X uma variável de *tratamento* e Z uma variável de desfecho, pode-se deduzir que conforme a primeira variável X causa a variável Y , a variável X determina o valor de Y . Logo, para que a segunda variável Y tenha um valor, ela depende do valor da primeira variável X . Da mesma forma, a segunda variável Y causa a terceira variável Z , determinando o valor final do desfecho Z e, assim, denotando que Z depende de Y para ter seu valor. Por fim, é possível concluir que Z e X são provavelmente dependentes entre si. Se a variável Z depende de Y e Y depende de X , provavelmente a variável Z depende de X para ter o seu valor final.^{121,134}

Porém, suponha que é observado um valor fixo na variável Y , como por exemplo $Y = a$. Quando valor de X é alterado, a variável U_y muda seu valor para compensar as alterações feitas por X , mantendo o valor de Y inalterado ($Y = a$). Como a relação de Z depende unicamente de Y e U_z no gráfico da Figura 9, o valor de Z se mantém inalterado. Desta forma é possível concluir que, se condicionar a

variável Y a um valor fixo, por mais que o valor de X se altere, Z não sofrerá alterações.¹²¹ Então, neste caso, uma nova relação para o gráfico da Figura 9 pode ser expressada como:

4. X é independente de Z , condicional a Y .

No caso apresentado pela Figura 10, é possível notar a variável Z como um *collider*. Sendo as variáveis exógenas U (U_x , U_y , U_z) independentes, as prováveis relações são:

1. X e Z são provavelmente dependentes;
2. Y e Z são provavelmente dependentes;
3. X e Y são independentes;
4. X e Y são provavelmente dependentes condicional a Z .

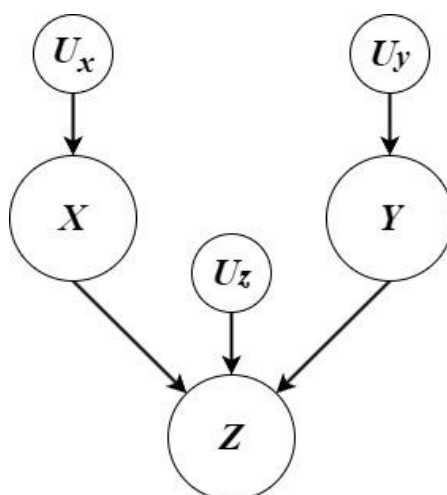


Figura 10. Exemplo de *collider*.

Fonte: O autor, 2024.

Associado a variável *collider*, os pontos 1 e 2 têm o mesmo comportamento. Ambas as variáveis X e Y estabelecem determinam o valor de Z . O ponto 3 enuncia a característica de que X e Y respondem apenas a U_x e U_y , respectivamente. Como U_x e U_y são independentes, não existe qualquer relação de causalidade que faça com que variações X interfiram no valor de Y , e vice-versa.¹²⁷

O ponto 4 se baseia no entendimento de que uma variável colisora (*“collider”*) quando condicional gera uma relação de dependência dos nodos pais. Ao ser

condicional, Z é filtrado, limitando as possibilidades de comparação de casos em que Z recebe o mesmo valor.¹²¹ Em um exemplo hipotético em que $Z = X + Y$, se Z for condicional a $Z = 6$ e for identificado que $X = 2$, o único valor possível para Y neste exemplo é de 4. Então fica evidente que, sabendo que o valor de Z recebe um valor, qualquer valor de X precisa ser compensado pela mudança do valor de Y . Isto denota uma relação de correlação entre as variáveis X e Y , sendo assim denominadas como *confundidoras* (“*confounding*”).¹³²

Por fim, a Figura 11 apresenta um gráfico causal onde a variável X é denominada como “*fork*”.¹²¹ Sendo as variáveis exógenas U (U_x , U_y , U_z) independentes, as possíveis dependências são:

1. X e Y são provavelmente dependentes;
2. X e Z são provavelmente dependentes;
3. Z e Y são provavelmente dependentes;
4. Y e Z são independentes, condicional a X .

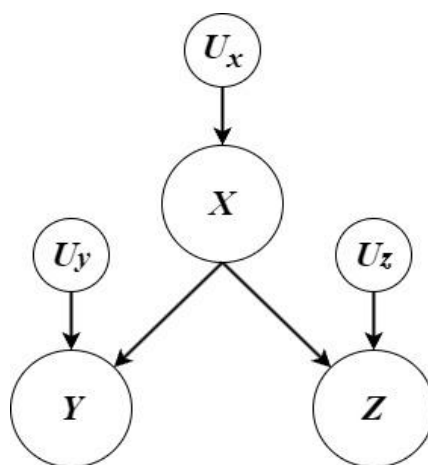


Figura 11. Exemplo de *fork*.

Fonte: O autor, 2024.

Os pontos 1 e 2 apresentam características de causalidade, onde X causa as variáveis Y e Z . Desta forma, como as variáveis Y e Z dependem do mesmo nodo X para ter o seu valor, é possível associar a probabilidade de dependência entre Y e Z . No ponto 4, quando se condiciona um valor a X (atribuindo-lhe um valor constante), a única forma de Y e Z mudarem é se U_y e U_z alterarem, respectivamente. Desta forma, é possível assumir a independência entre Y e Z quando condicionais a X .¹²¹

Estas relações e condições de dependência auxiliam a compreender melhor caminhos que estão bloqueados ou desbloqueados em um DAG. O condicionamento de uma variável, representado em um nodo, altera completamente como as relações do DAG se estabelecem, representando novas possibilidades de caminhos causais.^{6,106,122}

Comumente gráficos causais apresentam uma gama de relações com diversos caminhos possíveis entre os nodos. O processo de “*d-separation*” surge como método que permite analisar e prever as dependências compartilhadas entre os nodos do DAG.^{6,132} Dá a habilidade de informar não apenas se o modelo causal está errado, mas também informa onde está o erro. Esta qualidade permite ao pesquisador reparar falhas do modelo causal hipotético gerado ao invés de ter de recriar um modelo novo para suprimir eventuais incorreções.¹²¹

Em situações onde o modelo causal gerado é testado em outras populações, buscando um entendimento de um “escopo global”, o processo de *d-separation* se demonstra fundamental. Pelo fato de se basear apenas nas relações representadas em um DAG, o método de *d-separation* não precisa de funções para compreender relações causais ou probabilísticas entre as conexões dos nodos.¹³⁴

Na metodologia de *d-separation*, onde “*d*” significa “*direcional*”, se os nodos de um gráfico causal estão conectados por um caminho, estes nodos estão *d-connected*. Caso contrário, este caminho está *d-separated*. Desta forma, nodos que estão *d-connected* são provavelmente dependentes e nodos estão *d-separated* são independentes.^{129,141,143}

Conforme visto anteriormente, um caminho que possui *chains* e *forks* com uma variável intermediária condicionada estará bloqueado (Figuras 9, 11). Outra forma do caminho estar bloqueado é se existir um *collider* (Figura 10). Assim, o processo de *d-separation* elucidada o entendimento de que um caminho é bloqueado se ele estiver *d-separated*. Caso contrário, o caminho estará desbloqueado e será *d-connected*.¹³⁴

As representações de SCMs em DAGs é fundamental para compreender como as relações causais se elucidem. Porém, a representação em gráfico enuncia

incertezas de como que um grafo causal pode ser suficiente para responder a algum questionamento de causa a efeito. Dentre algumas soluções, uma ferramenta utilizada como teste são os critérios de *backdoor* e *frontdoor*.^{144,145}

Os critérios de *backdoor* e *frontdoor* são critérios aplicados a caminhos que determinam se uma trajetória entre duas variáveis representadas em nodos de um DAG possuem uma relação de causalidade. Caso a relação de causalidade não seja possível de ser inferida, isto é, caso o caminho possua relações bloqueadas (*d-separated*), um caminho denominado como caminho de *backdoor* poderá ter um conjunto de variáveis intermediárias condicionadas a fim de se viabilizar a relação de causalidade.¹²⁸

Um caminho que possui problemas para se identificar a relação de causalidade é denotado por possuir algum nodo que confunde o efeito que uma variável X tem sobre Y , por exemplo. Esta relação atribui a um nodo a característica de *confounding*. Um *confounding* se refere a um tipo de viés existente ao se estimar uma relação de causa e efeito.¹⁰⁴ Sendo *confounding* como o principal problema encontrado em estudos causais com dados observacionais, a influência desta característica introduz correlações espúrias (não causais), gerando um desentendimento da relação de causa e efeito de interesse.¹³¹

Existem vários tipos de *confounding* em estudos observacionais. Podem ser fatores biológicos, sociais, e genéticos, que podem ser de difícil mensuração. A forma como as associações e relações se estabelecem no DAG também podem implicar em *confounding*.¹⁰⁶

No caso de um gráfico causal, um caminho de *backdoor* representa um exemplo de onde o *confounding* pode ser identificado. O caminho de *backdoor* é um percurso de um nodo inicial X até outro nodo destino Y , onde a origem deste trajeto se inicia com uma flecha apontando para X . Esta relação perturbada entre X e Y pode ser *desconfundida* por se bloquear todo o caminho de *backdoor* que existir entre X e Y . Isto é, para desconfundir a relação entre duas variáveis X e Y , é necessário bloquear todo o caminho não causal entre as variáveis X e Y sem bloquear ou criar novos *confoundings* durante a trajetória da relação, preservando os caminhos direcionados entre as duas variáveis X e Y .⁴⁷

Na Figura 12 se tem a representação de um tratamento através de um antibiótico X no desfecho de Tempo de SLD Y . Também foram considerados para o modelo causal variáveis como peso W e a variável não mensurada renda familiar Z .

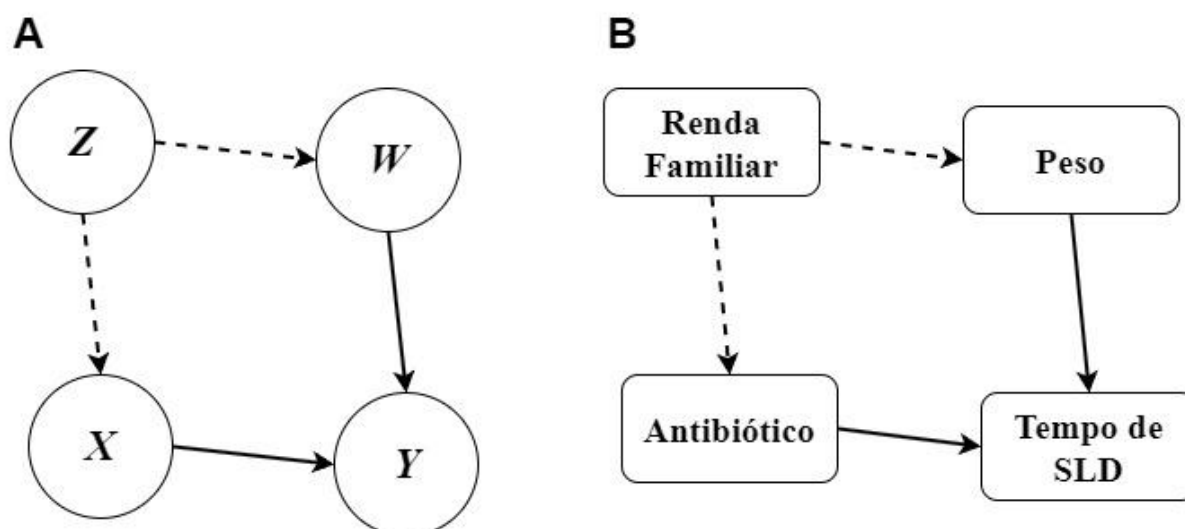


Figura 12. Exemplo de caminho de *backdoor*.

Fonte: O autor, 2024.

No gráfico causal representado pela Figura 12, existem alguns fatores a serem considerados. Como denotado, as variáveis X e W são causadoras do desfecho Y . O objetivo da análise causal é identificar a capacidade do tratamento X em causar a melhora do desfecho da doença Y . Ao se analisar os caminhos que levam ao desfecho Y é possível observar um caminho de *backdoor* entre as variáveis X , Z , W e Y ($X \leftarrow Z \rightarrow W \rightarrow Y$). A variável W neste caminho surge como *confounding*, pois confere os critérios de *backdoor*.¹²¹ Para que seja possível tratar este caminho, ajustando a variável W (tornando-a condicional e, assim, bloqueando o caminho) deve permitir a relação de efeito causal do tratamento X (antibiótico) no desfecho Y (tempo livre de doença).

Os efeitos de caminhos de *backdoor* implicam em problemas a serem resolvidos em diagramas causais. Todo caminho de *backdoor* deve ser bloqueado para permitir que outros caminhos possam transmitir suas relações de causalidade, caso contrário, o efeito causal será confundido (*confounding*). Em um trajeto onde X

tem influência causal em Y , condicionar caminhos de *backdoor* devem permitir que tal efeito aconteça.¹²⁸

Deve ser evitado condicionar nodos que sejam descendentes de X , visto que estes nodos filhos de X podem ter efeito de causalidade em Y . Caso tal fenômeno ocorresse, o caminho entre X e Y se tornaria bloqueado, distorcendo as associações causais de interesse. O mesmo pode acontecer caso se condicione as variáveis pai de X e Y .¹³¹

Da forma que um caminho de *backdoor* denota rotas espúrias, um caminho de *frontdoor* elucidada trajetos efetivamente causais. Em um percurso composto por um tratamento X em um desfecho Y , um caminho pode ser elucidado como caminho de *frontdoor* se um conjunto de variáveis Z interceptam todos os caminhos diretos de X para Y , se não houver caminhos de *backdoor* de X a Z e, por fim, todos os caminhos de *backdoor* de Z para Y tem de ser bloqueados por X .¹²¹

Um caminho de *frontdoor* em um DAG permite a identificação e compreensão da relação de efeito causal que uma variável X tem sobre outra Y , mesmo que um conjunto de variáveis Z que interceptam o caminho entre X e Y . O Teorema 1 de *Frontdoor Adjustment* demonstra como que tal efeito causal pode ser medido.¹³⁴ O método matemático usado no Teorema 1, que permite identificar todos os efeitos causais em um DAG, é denominado de *do-calculus*.¹²¹

Se Z satisfaz o critério de frontdoor relativos às variáveis X e Y e a probabilidade de X e Z é maior que 0, $P(x, z) > 0$, então o efeito causal de X em Y é identificado pela fórmula do Teorema 1 (considerando x' uma simples soma de índice):¹²¹

Teorema 1 (*Frontdoor Adjustment*):

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z)P(x')$$

Fonte: O autor, 2024.

4.2.4. *do-calculus*

As relações causais entre nodos em um diagrama causal podem ser expressadas (e beneficiadas) pelo método do *do-calculus*. O *do-calculus* é um conjunto de regras que fornecem um método sintático para verificar afirmações sobre intervenções e como elas interagem com desfechos de interesse. Podendo ser utilizado em dados observacionais, o *do-calculus* dá a um SCM a capacidade de determinar matematicamente quais as covariáveis (*confounding*) são apropriadas para serem tratadas, bem como, ajusta relações causais espúrias entre variáveis de interesse.⁶

O *do-operator*, operador matemático utilizado para denotar uma intervenção de uma variável X , corrige todas as suas arestas não causais que apontam para o tratamento X em questão. Esta técnica previne que o tratamento X propague relações não causais ao desfecho de interesse. A aplicação do método da randomização em estudos científicos exerce esta mesma qualidade.⁴⁷

Uma relação de probabilidade condicional entre duas variáveis X e Y , onde a probabilidade de $Y = y$ condicional a $X = x$, possui a fórmula probabilística representada por $P(Y = y|X = x)$. No caso de uma intervenção, a fórmula de probabilidade de $Y = y$ sofrer uma intervenção de forma que $X = x$ é representada por $P(Y = y|do(X = x))$.¹²¹

A representação $do(X = x)$ determina a relação de X como uma intervenção, propagando uma relação de causalidade em um SCM.⁶ A intervenção X tem de ser livre de *confounding*.¹⁰⁷

Inicialmente, em um gráfico causal com caminhos espúrios (não tratados), a probabilidade condicional entre X e Y é diferente da probabilidade de intervenção de X em Y : $P(Y = y|X = x) \neq P(Y = y|do(X = x))$. Uma expressão $P(Y = y|do(X = x), Z)$ altera um gráfico causal, removendo todas as arestas não causais que chegam a intervenção X . Esta modificação resulta em um gráfico onde a distribuição entre a probabilidade condicional e a intervenção se tornam a mesma: $P(Y = y|X = x) = P(Y = y|do(X = x))$. Tal fenômeno autoriza aos pesquisadores experimentarem os efeitos

de uma intervenção (similar ao que a randomização faz em estudo clínicos). Também permite a quem usa o método do *do-calculus* uma compreensão efetiva de quando relações entre variáveis são causais, e não apenas correlativas, em um DAG.¹²¹

A diferença entre intervir, através do operador $do()$, e condicionar uma variável consiste no fato de que a intervenção corrige seu valor. A intervenção altera a percepção que se tem de causalidade, enquanto o condicionamento não opera de tal forma. O condicionamento de uma variável filtra as possibilidades, reduzindo as opções do que pode ser interpretado e observado a partir das suas relações.^{47,146}

Dado um DAG G , com as variáveis de tratamento X , de desfecho Y e uma série de variáveis Z , onde Z é um conjunto de variáveis pai de X . A Fórmula 1, baseada no Teorema 1, calcula o efeito causal de X em Y , computando a associação de X e Y para cada valor z dos nodos pais Z e, em seguida, calcula a média desses valores. A equação esclarece como que os nodos Z que antecedem X (nodos pais de X) podem exercer influência sobre o seu valor, interferindo no valor preditivo da intervenção. Esta operação também pode ser interpretada como uma forma de ajustar ou controlar os nodos pai Z .^{121,147}

Fórmula 1A:

$$P(Y = y|do(X = x)) = \sum_Z P(Y = y|X = x, Z = z)P(Z = z)$$

Fonte: O autor, 2024.

Ou, alternativamente, a forma reduzida:

Fórmula 1B:

$$P(Y = y|do(X = x)) = \sum_Z \frac{P(X=x, Y=y, Z=z)}{P(X=x|Z=z)}$$

Fonte: O autor, 2024.

Em um exemplo hipotético representado pela Figura 13A, o tratamento por um antibiótico X ($X = 1$ ou $X = 0$) tem conexão com o desfecho de melhora da infecção (pneumonia) Y . Considere o nodo Z como um conjunto de covariáveis (como idade, peso, níveis de saturação, comorbidades, etc) que se relacionam e interferem no valor final de Y , e C (COVID-19) uma variável *confounding* com relações em X e Y . Se aplicar no tratamento X ($X = 1$) uma intervenção *do()*, a expressão utilizando o *do-operator* $P(Y | do(X = 1), Z)$ deve garantir que o efeito da intervenção não contenha efeitos de *confounding*.⁶ Para se poder avaliar os efeitos do tratamento ($X = 1$), em contraste com o não tratamento ($X = 0$), também se deve analisar a não-intervenção $P(Y | do(X = 0), Z)$. O efeito causal médio (*average causal effect, ACE*) pode ser calculado por: $P(Y | do(X = 1), Z) - P(Y | do(X = 0), Z)$.¹²¹

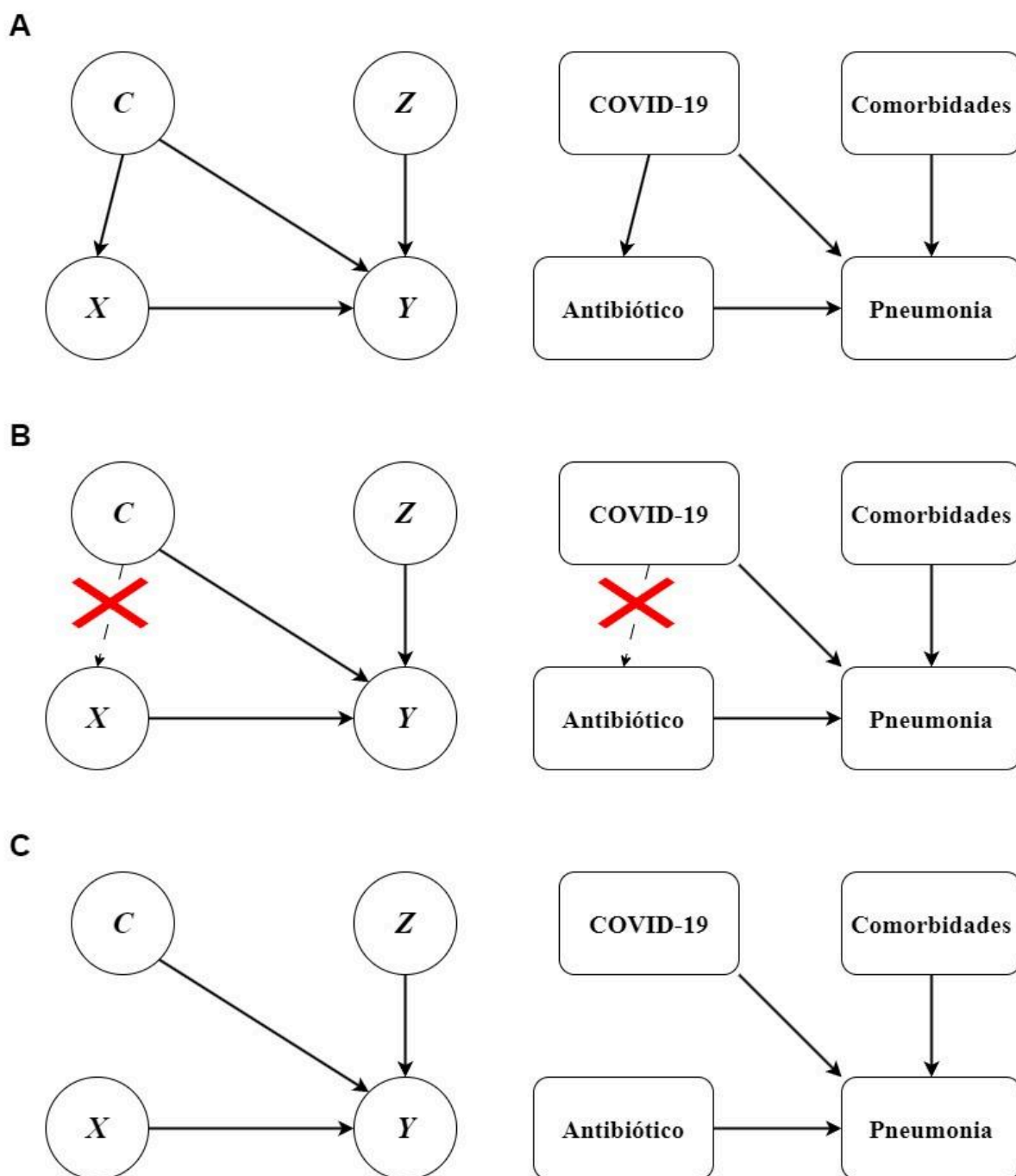


Figura 13. Exemplo hipotético de um DAG representando o tratamento de um antibiótico.
Fonte: O autor, 2024.

Na Figura 13B é possível observar a transformação do gráfico realizada pela aplicação do método do *do-calculus*. A aresta entre X e C é removida pelo *do-operator*, tornando a probabilidade de intervenção $P(Y | do(X = x), Z)$ a mesma

que a probabilidade condicional $P(Y | X = x, Z)$.^{106,121} Na Figura 13C é possível observar o novo gráfico gerado a partir do método.

Desta forma, para que seja possível ter uma compreensão de todas as qualidades do *do-calculus*, três regras se estabelecem:^{47,148}

1. Adição ou exclusão de observações:

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp Z|X, W)_{G_x^-}$$

Fonte: O autor, 2024.

2. Substituição de uma observação por uma intervenção, ou vice-versa:

$$P(y|do(z), w) = P(y|z, w) \quad \text{if } (Y \perp Z|W)_{G_z^-}$$

Fonte: O autor, 2024.

3. Adição ou exclusão de intervenções:

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if } (Y \perp Z|X, W)_{G_{XZ}^*}$$

Fonte: O autor, 2024.

Ou simplesmente:

$$P(y|do(x)) = P(y)$$

Fonte: O autor, 2024.

Considere um grupo de variáveis em um DAG G onde existem: o tratamento X , o desfecho Y e um conjunto de variáveis Z e W (Figura 14A). A Regra 1 diz que, quando uma variável Z é observada e ela é irrelevante ao desfecho Y , provavelmente as distribuições permanecerão inalteradas em Y .⁴⁷ Isto é, todas as arestas que apontam para X podem ser removidas contanto que Y seja independente de Z , dado X e W no “novo grafo” (Figura 14B). Se os nodos Y e Z

estão *d-separated* um do outro depois de analisar W e X no novo grafo (não existem conexões ou relações ativas entre W e X), então Z pode ser ignorado no gráfico da Figura 14B. Como o objetivo é calcular o efeito causal de X em Y , garantindo a que Y é independente de Z , dado X e W , a Regra 1 permite “ignorar Z ” de tal forma que $P(y|do(x), z, w) = P(y|do(x), w)$.¹⁴⁹

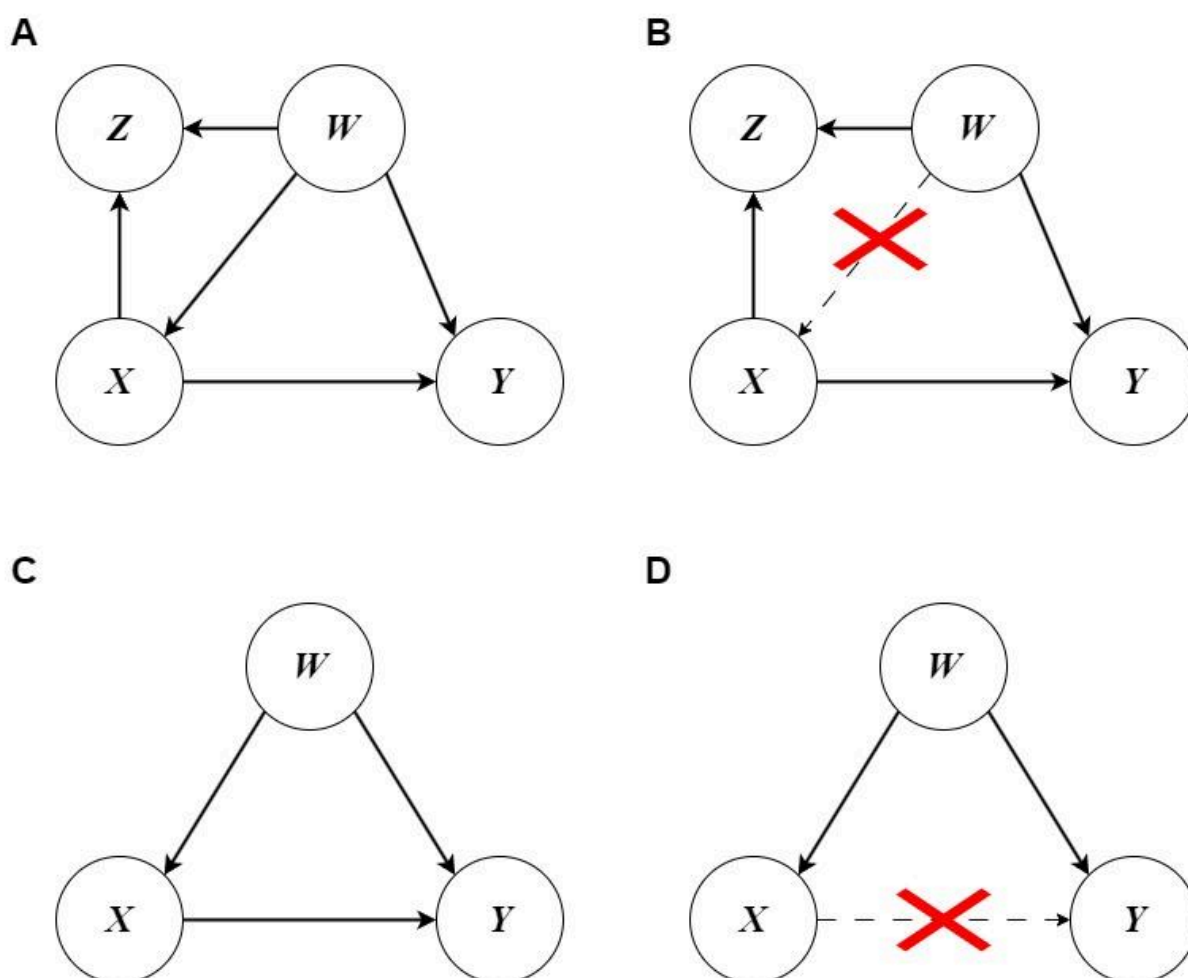


Figura 14. Exemplo aplicando as regras do *do-calculus*.

Fonte: O autor, 2024.

A Regra 2 se baseia no critério de *backdoor*. A Figura 14C apresenta o caminho de *backdoor* $X \leftarrow W \rightarrow Y$. Conforme o que é dito no critério de *backdoor*, W atua de forma a bloquear todos os caminhos de *backdoor* de X a Y se condicional em W . Isto quer dizer que, se W satisfaz o critério de *backdoor*, a Regra 2 pode ser aplicada de tal forma que $P(do(X)) = P(X)$. Neste caso, a aplicação da Regra 2 ajusta o grafo de tal forma que todas as flechas que saem de X são removidas, originando

um novo grafo onde Y é independente de X dado W (Figura 14D). As variáveis Y e X serão *d-separated* uma da outra após contabilizar W e as probabilidades condicionais e de intervenção serão igualadas: $P(Y | do(X), W) = P(Y | X, W)$.¹⁰⁷

Por fim, a Regra 3 indica que, se não existirem caminhos causais de X para Y , a intervenção $do(X)$ pode ser removida da fórmula. Esta regra indica que X é *d-separated* de Y e independentes entre si no novo gráfico, não sendo possível denotar associação entre X e Y . A Figura 14D mostra que não existem mais relações de causalidade (flechas ou arestas) de X para Y . Desta forma, a Regra 3 permite que $P(Y | do(X), W) = P(Y | W)$. Estas três regras são a origem de como o do critério de *frontdoor* se sucede.¹⁴⁸

4.4. Seleção de diagramas

Um modelo causal gerado a partir de dados observacionais comumente apresenta nuances específicas baseados no conjunto de dados do qual foi gerado. Estes dados provêm da coleta de estudos observacionais ou ensaios que foram realizados em coortes de uma população de origem. Características biológicas da população estudada podem implicar diretamente no comportamento do tratamento de doenças, alterando o curso das interpretações realizadas a partir de um estudo de causalidade. Também é comum observar experimentos biológicos que fazem uso de animais como substitutos aos humanos, por exemplo. As implicações que as características específicas que uma população pode ter representam um fator de grande relevância a ser considerado durante a confecção de SCMs.^{106,150}

Para a geração de um modelo causal é necessário se ter uma seleção de dados. Considerando que estes dados são provenientes de uma amostra de uma população, esta relação amostral pode implicar em não representar adequadamente a população, confundindo as relações de causalidade de um modelo gerado a partir destes dados. Estes vieses de estudo que não são associados diretamente a presença de causas comuns de tratamento e desfecho são denominadas como *selection bias*.¹⁴⁸

Selection bias é um conceito que compreende uma série de vieses de estudo que se originam da amostra de dados selecionada para análise. A amostra de dados selecionada pode conter artefatos ou características que, caso um SCM seja gerado a partir desses dados, este modelo causal pode não funcionar se testado em outra população.¹⁵¹

Em um gráfico causal, representações podem demonstrar a existência de *selection bias* e, também, existem mecanismos que podem ser feitos para tratar tais vieses. Na Figura 15, supondo um tratamento X que causa o desfecho Y , um nodo especial S surge como um indicador de seleção. Este indicador S não é uma variável e funciona de tal forma que as suas arestas denotam potenciais mudanças populacionais (ou amostrais) em variáveis com as quais se relaciona.¹⁰⁶

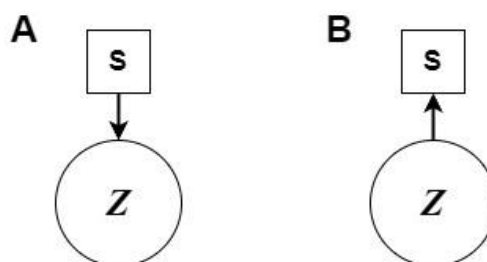


Figura 15. Representação do nodo de seleção.

Fonte: O autor, 2024.

A Figura 15A denota uma flecha apontando de S para Z . Isto indica que o valor de Z pode alterar caso testado em outra população. A Figura 15B contém uma conexão onde existe uma flecha apontando de Z para S . Esta relação implica na potencial presença de um *selection bias*, isto é, há diferenças que não são entre populações mas sim entre domínios de amostra de dados.¹⁰⁷

Recuperar um modelo causal com um *selection bias* é uma tarefa de complexa resolução.⁴⁷ Em um DAG, é possível se ajustar para incluir o indicador de seleção, condicionar caminhos usando as regras de *d-separation*, ou até mesmo bloqueando caminhos de *backdoor* que introduzem variáveis com indicador de seleção. Porém, há casos em que nem sempre é possível realizar tais operações. Dentre os métodos para se tentar minimizar ou solucionar *selection bias* estão algoritmos propostos por Elias Bareinboim, com vistas a estratificar a amostra de

dados, completar dados faltantes com dados externos e até mesmo com a aplicação de um novo ECR, quando possível.^{106,152–155}

Um indicador de seleção S pode também demonstrar quando tratamentos e variáveis mediadoras (*mediators*) podem diferir se o modelo for testado entre outras populações.¹⁰⁷ A Figura 16A apresenta um nodo de tratamento X , um nodo de desfecho Y e um nodo de seleção S . Enquanto a Figura 16B apresenta um nodo de tratamento X , um nodo Z como um mediador, um nodo de desfecho Y e um nodo de seleção S que denota que tais características mudam entre populações.

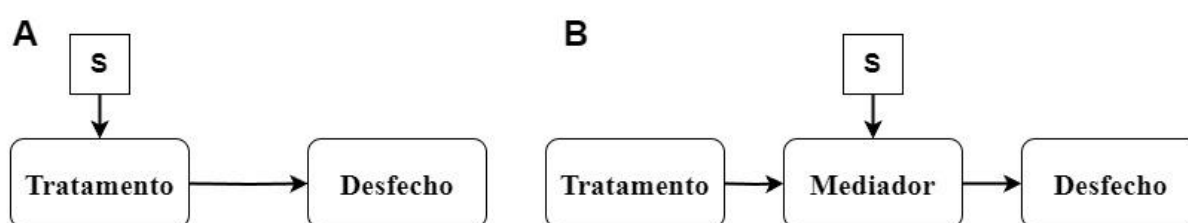


Figura 16. Exemplo de DAGs contendo um nodo de seleção.

Fonte: O autor, 2024.

Em um exemplo hipotético onde um mesmo ECR é aplicado na Finlândia e no Japão, seguindo as mesmas normas e tratamentos, um gráfico causal pode conter uma variável “raça” sendo apontado por um nodo de seleção S para se denotar as mudanças populacionais (Figura 17A). Outro exemplo (Figura 17B) pode ser representado pela existência de “biomarcadores” como causadores de um desfecho. Neste caso, um estudo tem biomarcadores positivos e negativos de pacientes, enquanto outro estudo possui apenas pacientes com biomarcadores positivos. Se em um caso onde for analisado um tratamento por um medicamento X , e a conexão for mediada por um nodo de “doses do medicamento” a um desfecho Y , um nodo de seleção S deve indicar a possibilidade que em um estudo incluiu 1 e 2 doses do medicamento, enquanto outro estudo incluiu 3 e 4 doses (Figura 17C). Também pode existir casos como o do exemplo em que um estudo possui informações de exames laboratoriais enquanto outros não (Figura 17D).

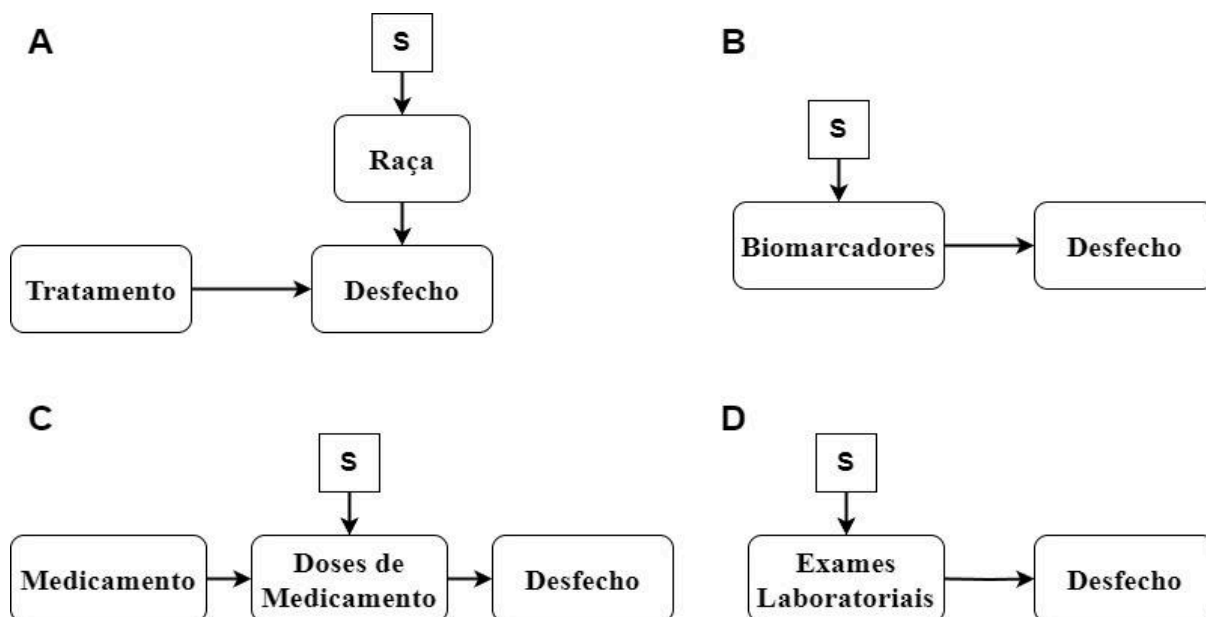


Figura 17. Exemplos contendo um nodo de seleção.

Fonte: O autor, 2024.

Um *confounding* presente em um caminho implica em uma relação de distorções entre as associações de variáveis, produz vieses de estudo e enuncia correlações não causais entre um tratamento e um desfecho. Porém, tais caminhos espúrios podem ser tratados de forma a viabilizar relações de causalidade.¹²⁸

Dentre as características já mencionadas: *colliders*, caminhos de *backdoor*, *selection bias*, variáveis não medidas e até mesmo fatores desconhecidos podem produzir *confounding*, gerando vieses de estudo. Existem outras formas em que estes vieses podem se manifestar.¹⁰⁶

Em uma situação em que um *collider* está presente, o caminho está bloqueado. Para que seja possível viabilizar a relação de causalidade, condicionar a variável que apresenta o *collider* “abre” o caminho, desbloqueando a variável *collider*. Em caminhos de *backdoor*, os métodos foram demonstrados. Dentre as técnicas, se bloquear todo o caminho de *backdoor* existente isso deve viabilizar a relação de causalidade entre um tratamento e um desfecho.¹²¹

Estas duas técnicas se utilizam de condicionar variáveis em um caminho para ajustar o efeito. Um *collider* ao ser condicionado desbloqueia o caminho. Da mesma forma, um caminho de *backdoor* desbloqueado pode transmitir uma associação não

causal. Se a variável eleita a ser condicional for controlada, o caminho de *backdoor* se tornará bloqueado, permitindo a relação de causalidade.¹²¹

Apesar da técnica de condicionar surgir como um método viável para ajustar relações em um DAG, o condicionamento de variáveis pode gerar vieses. Se uma variável *collider* for controlada, implica em que o entendimento que se tinha dela se torna distorcido, isto implica na possibilidade de alterar a estimativa final do efeito, ainda, podendo introduzir uma associação negativa (ou inversa) onde não existia efeito causal. A ocorrência de correlação negativa espúria se manifesta somente quando o resultado combinado das duas variáveis de interesse é controlado. Isto gera a impressão de que, para que uma associação de causalidade se estabeleça, ela precisa ser “compensada”. Desta forma, também se deve evitar controlar descendentes de um *collider* com tal característica.¹³¹

A tarefa de mensurar o erro gerado por um *confounding* é complexa, de precisão inexata.^{47,106,152,153} De tal forma, para evitar os vieses de difícil mensuração produzidos por um *collider*, primeiramente se deve compreender a localização no DAG onde o *collider* ocorre e, se possível, executar análises que não sejam condicionais (antes de simplesmente condicioná-las). Caso contrário, o *collider* pode gerar compreensões inconclusivas sobre associações de causalidade.^{47,128,131}

A falta de informações em um conjunto de dados retrospectivos também é um dos principais motivos para ocorrência de *confounding* e vieses de estudos de inferência causal. O nome destes enviesamentos são denominados de *nonresponse* e *attrition*.^{106,156}

O *nonresponse bias* é um tipo de viés que pode representar a perda de informações nos dados, impactando o desfecho, por quaisquer motivos. Por exemplo, um paciente relutante em fornecer informações ou que omite questões importantes, por motivos não relatados a quem registra os dados, pode gerar um *nonresponse bias*.^{106,156}

O *attrition bias* ocorre quando a coorte de dados de um estudo representa apenas indivíduos com uma característica comum, esta coorte também pode ocorrer por motivos aleatórios. Por exemplo, considere na Figura 18 um estudo longitudinal

onde se analisa o uso de um medicamento X para tratar sequelas de pacientes que tiveram um quadro de pneumonia grave provocado e complicado pela COVID-19. Sendo Y um cenário onde há melhoria ou não dos sintomas provocados pelas sequelas da doença, considere U um conjunto de variáveis associadas à imunossupressão, de difícil mensuração, e P problemas respiratórios prévios (que antecedem o quadro grave de pneumonia). Em um cenário onde os participantes deste estudo que apresentaram problemas respiratórios P venham a falecer por complicações prévias ou apenas solicitem desistência do estudo por conta de terem questões de saúde mais sensíveis, a variável P se tornará desbalanceada. Isso gera vieses de estudo denominado como *healthy worker bias* (ou viés do participante saudável). Neste sentido, os fatores de imunossupressão U e problemas respiratórios prévios P ainda terão atrito com o possível desfecho do tratamento do medicamento X no cenário de melhoria dos sintomas Y . Este atrito pode levantar caminhos de *backdoor*, gerando associações espúrias entre as variáveis de interesse (Figura 18).^{106,156}

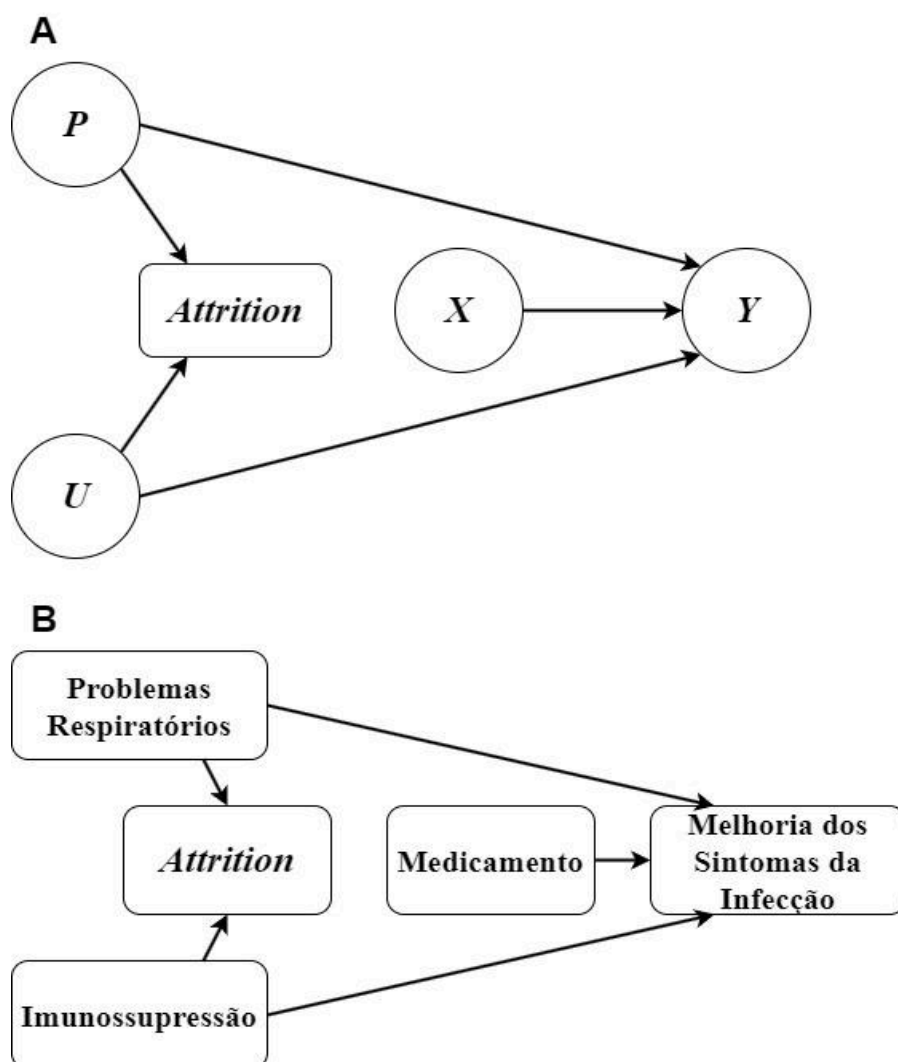


Figura 18. Exemplo de um estudo longitudinal para se analisar o tratamento do medicamento X em um desfecho pneumonia.

Fonte: O autor, 2024.

4.2.5. Seleção de diagramas

Os métodos de inferência causal apresentados possuem diversas qualidades, conforme mencionado. Dentre elas, independente das equações existentes nos modelos em questão, os métodos de *d-separation* e os critérios de *frontdoor* e *backdoor* permitem ao pesquisador interpretar as associações existentes entre duas variáveis, determinando sua relação de causalidade com um desfecho. Em muitas

ocasiões, acaba não sendo necessário identificar toda a relação de efeito causal entre todos os parâmetros do modelo.^{121,141,157}

Estimar os efeitos diretos e totais de causalidade através de estudos não experimentais é uma tarefa complexa. Determinar se uma relação é “suficientemente causal” em um estudo observacional exige que um conjunto mínimo de condições e eventos seja satisfeito para produzir um efeito (a um desfecho). Tal questão é denominada como “*identifiability*” e equivale a representar os efeitos diretos e totais de coeficientes de um conjunto mínimo de variáveis em um caminho através das covariâncias ou coeficientes de regressão.¹²⁰

O entendimento de uma relação “suficientemente causal” é comum nos estudos de causalidade. Indagações realizadas às relações totais e diretas de causalidade surgem de tal forma a buscar alternativas ao acontecimento dos fatos.^{128,134,144,158} Tais eventos são dúvidas manifestadas através de questionamentos contrafactuais com o objetivo de buscar um entendimento mais amplo sobre diagnósticos e prognósticos. Estas perguntas sobre os processos enunciam possibilidades que não necessariamente aconteceram.^{159,160}

Em estudos experimentais, apenas um dos tratamentos (tratamento ou placebo) é possível ser dado a um paciente. Em uma situação real, não existem duas cópias do mesmo paciente para que seja possível de ser observado todos os efeitos possíveis nos desfechos promovidos pelo tratamento e pelo não tratamento.¹³¹

Este conceito encerra o escopo denominado como “*Ladder of Causation*”. O primeiro degrau estabelece as probabilidades condicionais para prever um evento com base nas associações entre variáveis. O segundo degrau envolve determinar relações causais através de intervenções. E por fim, o terceiro degrau engloba a possibilidade de escolha entre tratamentos para indivíduos. O questionamento contrafactual “*E se?*” permite comparar desfechos sob as mesmas condições, diferenciando-se apenas um aspecto em dados observacionais. Este aspecto em questão é denominado como “*antecedent*”.^{47,121}

4.2.6. Inferência Causal e Contrafatos

Os métodos de *d-separation* e critério de *backdoor* funcionam adequadamente para identificar onde relações devem ser ajustadas e tratadas, porém não permitem que sejam feitas suposições e implicações sobre os relacionamentos das variáveis. Sistemas Lineares reforçam os métodos aplicados em modelos não paramétricos e, também, auxiliam a entender os efeitos de inferência causal entre variáveis de interesse.¹⁶¹⁻¹⁶³

As suposições realizadas puramente em gráficos causais, sem análise de qualquer tipo de equação associada, se baseiam em pressuposições de que as relações entre as variáveis são lineares e que todos os erros possuem distribuições normais. Estas presunções baseadas em distribuições normais podem ser denominadas como “expectativas”. Quando aplicadas em sistemas lineares, possuem as propriedades de representação eficiente, substituibilidade de expectativas por probabilidades, linearidade das expectativas e invariância dos coeficientes de regressão.¹⁶⁴

Uma distribuição normal multivariada, por ser baseada em expectativas, permite ao pesquisador o poder de expressar probabilidades condicionais desconhecidas através de expectativas condicionais. Isto é, mesmo não existindo uma “tabela de probabilidades” para as variáveis discretas, as técnicas de independência condicional que moldam as relações dos gráficos podem ser expressas em formato de relações de igualdade entre expectativas condicionais. A característica de substituir expectativas por probabilidades enuncia a possibilidade de uso de métodos de regressão para determinar informações de causalidade.¹²¹

A expressão condicional de independência de Y e X , dado Z (um conjunto de variáveis), é dada por $P(Y|X, Z) = P(Y|Z)$. Desta forma, a expressão das expectativas pode ser escrita por $E[Y|X, Z] = E[Y|Z]$.⁶

Uma variável condicionada, neste caso, é denominada regressor e é uma variável inserida para regressão. Cada expectativa condicional $E[Y|X_1, X_2, \dots, X_n]$ é dada pela combinação de variáveis condicionadas (Fórmula 2).^{162,163}

Fórmula 2 (expectativas condicionais):

$$E[Y|X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] = r_0 + r_1 x_1 + r_2 x_2 + \dots + r_n x_n$$

O coeficiente parcial de regressão é denominado como inclinação e essas inclinações são representadas por $R = r_1, r_2, \dots, r_n$. A magnitude dessas inclinações não depende das variáveis condicionais (regressores), apenas dependem de quais variáveis são eleitas para regressão. Esta qualidade elucida que, na Fórmula 2, a sensibilidade de Y depende apenas da seleção de variáveis feita, e não dos valores das variáveis $X_i = x_i$.^{6,121,165}

Diferentemente dos coeficientes de regressão R , os coeficientes de caminho β (ou coeficientes estruturais) permitem especificar funções no modelo através de coeficientes representados ao longo das arestas de um gráfico causal. Este coeficiente β quantifica a relação de efeito direto entre arestas de X para Y ($X \rightarrow Y$), denotando no modelo a contribuição que a variável independente X tem na função que define a variável dependente Y . Isto quer dizer que o coeficiente de caminho indica uma relação de causalidade, enquanto um coeficiente de regressão indica apenas a correlação entre variáveis (relação não necessariamente causal).¹⁶²

Para um conjunto de três variáveis X , Y e Z , sendo Z uma variável intermediária (*mediator*) entre X e Y , a fórmula de efeito controlado direto (*controlled direct effect, CDE*) de Y alterando o valor de $X = x$ para $X = x'$ é definida por:¹²¹

Fórmula 3 (Controlled Direct Effect):

$$CDE = P(Y = y|do(X = x), do(Z = z)) - P(Y = y|do(X = x'), do(Z = z))$$

Agora considere o exemplo a seguir onde seja de interesse estimar o efeito direto de Z em Y contendo as equações estruturais:¹²¹

$$X = U_X$$

$$Z = aX + U_Z$$

$$W = bX + cZ + U_W$$

$$Y = dZ + eW + U_Y$$

Como constatado nas equações, W é o único pai de Y, fazendo a reescrita da equação aplicando-se a Fórmula Y na forma das expectativas, se obtém:¹²¹

$$DE = E[Y|do(Z = z + 1), do(W = w)] - E[Y|do(Z = z), do(W = w)]$$

Reduzindo a equação, considere os seguintes passos para constatação do efeito direto (DE):

$$E[Y|do(Z = z + 1), do(W = w)] = d(z + 1) + ew$$

$$E[Y|do(Z = z), do(W = w)] = dz + ew$$

$$DE = d(z + 1) + ew - (dz + ew)$$

$$DE = d$$

A aplicação dos *do-operators* remove adequadamente as equações do modelo, sem necessitar fazer suposições sobre correlações entre erros ou fatores desconhecidos U . O resultado da fórmula de efeito direto (*direct effect, DE*) resulta no coeficiente de caminho entre Z e Y ($DE = d$).¹⁶⁴

Estes conceitos assumem que os efeitos diretos são, de alguma forma, conhecidos a priori (estimados em intervenções experimentais, por exemplo). Estimar os efeitos totais e diretos em dados não experimentais é uma questão denominada como "*identifiability*".¹⁶⁶ O problema de *identifiability* consiste na dificuldade que se existe em expressar os efeitos totais e seus respectivos coeficientes de caminho em termos das covariâncias σ_{XY} ou dos coeficientes de regressão $R_{YX.Z}$, onde X e Y são duas variáveis de interesse para se calcular os efeitos causais e Z é um conjunto de variáveis que satisfaçam o critério de *backdoor* no modelo.¹⁶⁷

Em uma situação em que não se tem todas as equações e parâmetros estruturais do modelo, mas se possui o DAG correspondente, é possível identificar os efeitos de causalidade. O critério de *backdoor* enuncia que um conjunto de variáveis Z precisam ser ajustadas para determinar o efeito causal de X em Y . Desta forma, é possível estimar a expectativa condicional de Y dados X e Z . Usando a dependência resultante entre X e Y e o valor médio de Z é possível calcular o efeito que X tem em Y . A regressão em Z adiciona as variáveis que satisfazem o critério de *backdoor* na equação, fazendo com que ocorra o bloqueio dos caminhos espúrios de X e Y e a prevenção de informações incorretas no coeficiente de X .^{121,168}

Isto é, em outras palavras, o procedimento efetua os seguintes passos:¹⁶⁸

1. Realiza-se a busca no modelo por um conjunto de covariáveis Z que satisfaçam o critério de *backdoor* para as variáveis de tratamento X e desfecho Y .
2. Após, aplica-se o método de regressão para regredir Y em X e Z .
3. O coeficiente associado a X no resultado da equação indica a relação de efeito causal de X em Y .

No exemplo representado pela Figura 19 o grafo representa um sistema linear e é de interesse calcular os efeitos causais que X tem em Y . Utilizando o critério de *backdoor*, o caminho $T \rightarrow Y$ elucidada a necessidade de se ajustar T para viabilizar a relação de causalidade de X para Y . Desta forma, a estrutura do gráfico autoriza, após o ajuste em T , que W seja ignorado. Isso permite a regressão de Y em X e T através da equação de regressão $y = r_X X + r_T T + \epsilon$. O coeficiente r_X representa o efeito total de X em Y e ϵ um coeficiente de erro.¹²¹

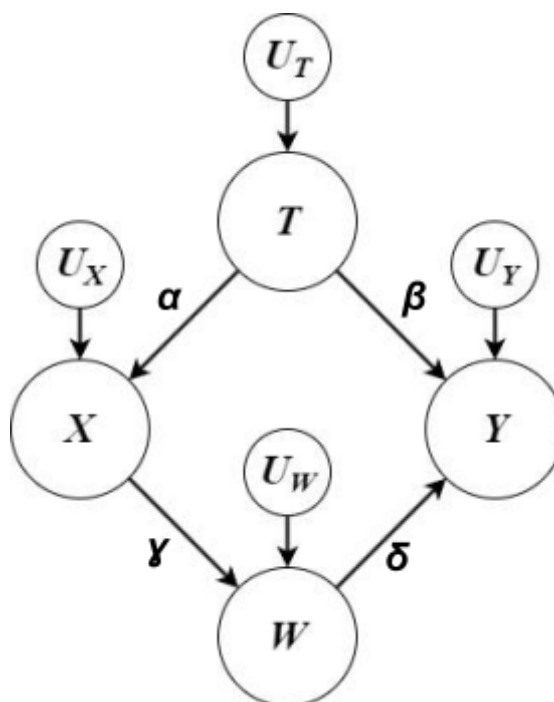


Figura 19. Exemplo para se analisar os efeitos de X em Y em um sistema linear.

Fonte: O autor, 2024.

Agora suponha que seja de interesse de se calcular a relação de efeito direto de X em Y . O coeficiente estrutural α denota a relação de efeito direto e, em um sistema linear, a função que define Y é dada por: $y = \alpha x + \beta z + \dots + U_Y$.¹⁶³

Considere a Figura 20 o DAG G que deriva do gráfico da Figura 19, com a presença do coeficiente estrutural α que denota a relação de efeito direto de X e Y . Neste caso, a Figura 19 não possui uma aresta direta conectando X a Y , logo o coeficiente estrutural é $\alpha = 0$. Isto indica que o efeito direto que X tem em Y é igual a 0 .¹²¹

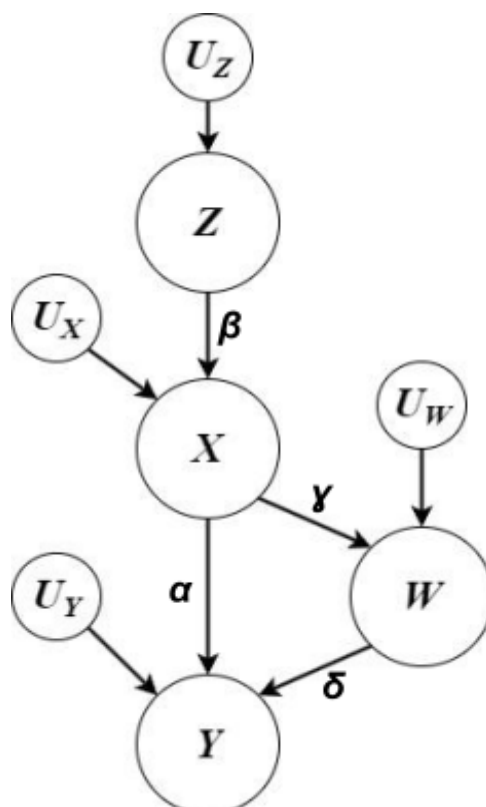


Figura 20. Exemplo derivado para se analisar os efeitos de X em Y em um sistema linear.

Fonte: O autor, 2024.

A magnitude de α ainda pode ser identificada mesmo que o modelo não determine seu valor. A “Regra de Regressão para Identificação” fornece um método similar ao processo do critério de *backdoor*, determinando quais variáveis devem constar na equação e que qualquer parâmetro (α) pode ser identificado pela regressão de mínimos quadrados ordinários (*ordinary least square, OLS*).¹⁶²

A “Regra de Regressão para Identificação” permite que o efeito direto de X em Y seja identificado no grafo G na Figura 20:¹²¹

1. Remove-se as arestas de X a Y, caso existirem. Isso resulta em um grafo G_α (Figura 21).
2. Se existirem um conjunto de variáveis Z que façam com que X e Y fiquem *d-separated*, deve-se regredir Y em X e Z. Na Figura 21 é possível observar que W atua de tal forma a tornar X e Y são *d-separated*. Então, efetua-se a regressão de Y em X e W usando-se a equação de regressão $Y = r_X X + r_W W + \epsilon$.

3. O coeficiente r_x na equação resultante é igual ao coeficiente estrutural α , isto é, o valor da relação efeito direto de X em Y .

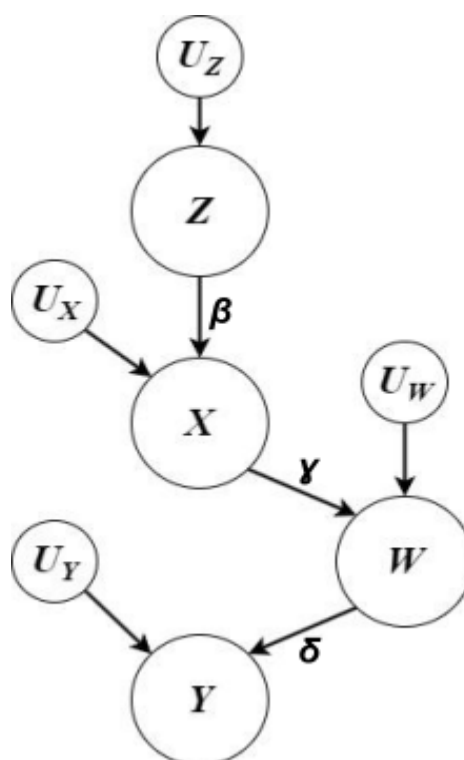


Figura 21. Exemplo denotando os efeitos de X em Y em um sistema linear.

Fonte: O autor, 2024.

4.3. IRAS

As IRAS estão entre os principais indicadores de óbito hospitalar nos Estados Unidos da América e no Brasil.^{51,55} Alta taxa de morbidade, aumento da resistência de microrganismos aos antimicrobianos, prolongamento de estadia em leito hospitalar e aumento de custos para tratamento de pacientes internados estão entre os fatores complicadores para quem enfrenta tal contexto.⁴⁸ Por definição, as IRAS são infecções que ocorrem 48 horas após a admissão hospitalar do paciente ou após a alta, quando estiver relacionada a internações anteriores ou a procedimentos hospitalares prévios.⁴⁹

Com a necessidade de se buscar um entendimento melhor do impacto das IRAS no contexto da área da saúde, grupos buscaram formas de estudar medidas

que pudessem estar associadas à prevenção, diagnóstico, causa, ou até mesmo risco elevado para que tais infecções surgissem.⁴⁸ Estes grupos consolidados foram denominados como CCIHs. Estas comissões são responsáveis pela vigilância epidemiológica e realizam o “controle de infecções” através da busca ativa das IRAS, de forma sistemática e contínua, em clínicas e unidades hospitalares. Estima-se que um terço das infecções hospitalares podem ser prevenidas pela realização do trabalho dos profissionais de uma CCIH.⁵⁰

O processo de busca ativa no controle de infecções é realizado com o fim de reconhecer eventos e condições que influenciam no surgimento e na ocorrência das infecções em ambientes clínicos.⁵⁷ Este processo segue critérios explanados por manuais (NHSN e Anvisa), que funcionam como ferramentas para orientar profissionais a identificar e prevenir infecções. Estes manuais orientam boas práticas e sugerem que a ocorrência das IRAS é associada a eventos que indiquem a presença de infecção em um paciente através. Estes eventos podem ser sintomas, exames laboratoriais, procedimentos invasivos, uso de antibióticos ou tempo de estadia no hospital. Outros fatores associados com o surgimento e gravidade das IRAS podem ser denominados como fatores preventivos e fatores de risco.^{49,169-171}

Um fator de risco corresponde a aspectos pessoais de saúde de um paciente. Podem ser condições associadas ao ambiente em que se encontra, características inatas ou herdadas que estão associadas diretamente à saúde do enfermo. Já um fator preventivo (ou protetivo) corresponde aos mesmos aspectos pessoais de saúde do paciente, sendo elas associadas ao ambiente, e características inatas ou herdadas, porém com relação à prevenção ou mitigação de condições importantes de saúde a serem prevenidas.¹⁷²⁻¹⁷⁴

Estes fatores de risco e preventivos são estudados na área da saúde através de estudos observacionais para elucidar associações entre variáveis e o desfecho de interesse. As associações não possuem relações de causalidade efetivamente comprovadas, visto que esta abordagem realizada pelos estudos é puramente correlacional e, desta forma, é fundamentalmente essencial considerar *confounding*.^{107,109,175} Mesmo sendo correlações, as variáveis contendo tais associações de risco foram estudadas arduamente pelos pesquisadores clínicos de

tal forma a servir precursor dos estudos de causalidade e base teórica para a construção de gráficos causais.¹⁷⁶

Os efeitos causais de fatores de risco e fatores preventivos nem sempre podem ser transpostos em uma intervenção. Além de que os efeitos causais destes fatores comumente são diferentes dos efeitos causais das intervenções, a remoção ou ajuste de um destes fatores pode não implicar em normalização ou viabilização do efeito causal.¹⁷⁷

A seleção de variáveis a serem colocadas em um modelo causal é uma tarefa complexa e, quando feita de forma inadequada, pode representar em viés de estudo.¹⁰⁶ A seleção de um “conjunto suficiente de fatores” que devem constar no modelo é fundamental para apoiar as suposições causais.¹⁶⁸ Estas variáveis devem seguir a narrativa de “história causal”, onde o conjunto de variáveis presentes satisfaz as hipóteses e suposições do estudo, e satisfazem o critério de *backdoor*.^{6,105} A seleção das variáveis de estudo e relações entre os nodos do DAG, idealmente, deve ocorrer antes de uma análise mais profunda de dados observacionais. Isso permite estabelecer adequadamente as suposições e hipóteses previamente elaboradas de forma a não se criarem vieses provocados pelos dados ou pelas características populacionais.^{47,106,121} Porém, tal afirmação não é impeditiva para que o uso de algoritmos de regressão ou “*feature importance*” possam ser usados a fim de auxiliar a identificar variáveis, sejam elas não observadas ou inadequadas de estarem presentes no modelo (variáveis de alta correlação)^{178–181}.

Desta forma, entender como as associações destes fatores se comportam quando estudadas em teses causais é fundamental para elucidar as causas de meras correlações, quando possível.⁴⁷ Seguindo o exemplo de Infecção de Sítio Cirúrgico (ISC), existem critérios, fatores preventivos e fatores de risco que podem ser inseridos como variáveis de estudos de causalidade. Índices e *scores* foram criados a partir destes fatores para estabelecer uma metodologia para a gravidade de um paciente e seus riscos de contrair uma infecção por conta do procedimento cirúrgico. Estes índices e *scores* sofreram diversas alterações ao longo dos tempos visto que estabelecem apenas critérios correlativos.^{182–185}

Sendo estes fatores sugestões de variáveis de estudo para teses de causalidade, os critérios associados ao paciente são comorbidades modificáveis e tratáveis como diabetes, imunossupressão, obesidade, neoplasia, hipertensão arterial, e outras doenças significativas. O uso de tabaco é uma condição associada ao paciente que pode ser modificada. Idade avançada, história de infecções prévias de pele e tecidos moles e histórico de radioterapia são fatores associados ao paciente não modificáveis. Alterações nos parâmetros de exames laboratoriais como hemogramas, leucócitos, glicemia e albumina podem enunciar questões associadas ao paciente que indiquem um maior risco de contrair uma ISC.^{182,183,185}

Dos eventos associados ao procedimento cirúrgico, podem ser considerados os fatores de contaminação do ar da sala cirúrgica, uso de anticoagulantes, transfusões de sangue, oxigenação, tempo de cirurgia (depende do tipo de procedimento), hipotermia perioperatória, hiperglicemia pós operatória, cuidados com a ferida, técnica de procedimento cirúrgico (associado a equipe cirúrgica), contaminação da ferida pela flora na pele do paciente, contaminação da ferida através dos instrumentos cirúrgicos e contaminação da ferida através das roupas, vestimentas da equipe cirúrgica, sapatos e outros utensílios ou vestimentas. Alguns estudos trazem outros fatores como por exemplo de que os pelos só devem ser removidos se estiverem em excesso; uso de descolonização intranasal e antisséptico de pele na região do procedimento cirúrgico; profilaxia antimicrobiana dentro de 1 hora após a incisão; taxa de adequação de checklists de boas práticas para prevenção de ISC.¹⁸²

5. RESULTADOS: Aplicação do Framework de Inferência Causal

A sessão aborda a aplicação do *framework* usando-se como exemplo o caso de Febre Puerperal, estudado por Ignaz Semmelweis no século 19.

5.1. O caso da Febre Puerperal

Durante o século 19, a Febre Puerperal era uma condição devastadora responsável por uma alta taxa de mortalidade entre mulheres europeias em período pós-parto. Seus sintomas mais comuns estão entre febre alta, dor abdominal e mau odor das regiões genitais. Em casos mais graves os sintomas variam de inchaço ou dor nas extremidades ou articulações, náusea e vômito, corrimento vaginal copioso, dor de cabeça, mialgias, dispneia, irritação na pele, faringite, confusão e mal-estar geral, levando o paciente para uma rápida progressão de óbito.^{186–188}

No final da década de 1830, o número crescente de óbitos por Febre Puerperal estava gerando muita preocupação aos médicos da época. Ignaz Semmelweis foi um médico húngaro responsável por fazer os primeiros estudos de causalidade sobre o assunto (Figura 22). Ele fez uso de um raciocínio que aprendera na “Escola de Medicina de Viena” para investigar as possibilidades de causa para tantos casos de Febre Puerperal. Tal método procedural, baseado em critérios de exclusão estritos do processo de “*modus tollendo ponens*”, permitiu a Semmelweis e seus discípulos a compreenderem a possível causa para a doença através de práticas e dados estatísticos básicos das duas clínicas da época.¹⁸⁸

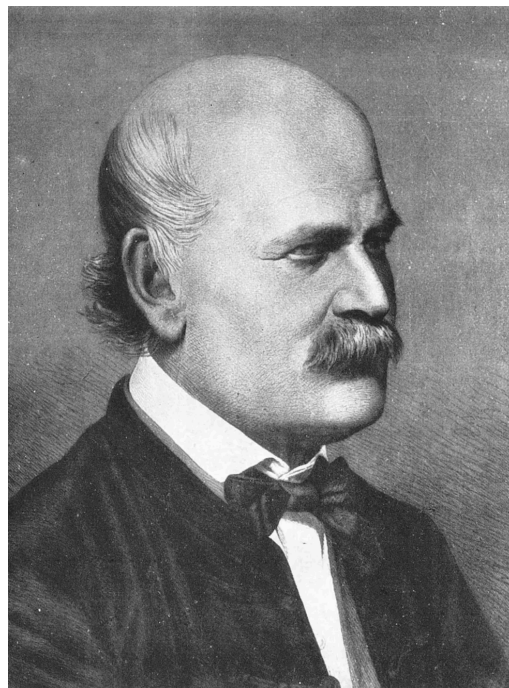


Figura 22. Retrato de Ignaz Semmelweis por Eugen Doby.¹⁸⁹

Fonte: Desconhecido, Domínio Público.

Uma clínica obstétrica da região foi separada em duas unidades em 1839 e ao longo dos anos seus dados foram estudados e observados por Semmelweis. Enquanto uma clínica foi alocada para a instrução de novos médicos obstétricos, onde realizavam diversos estudos anatômicos e autópsias em cadáveres, a outra ficou apenas com parteiras. A clínica representada pelos médicos obstétricos registrava um número muito maior de óbitos relacionados a Febre Puerperal que a clínica composta apenas por parteiras.¹⁸⁷

Desta forma, após assumir um cargo na Clínica Obstétrica do “Hospital Geral de Vienna” em 1º Julho de 1846, Semmelweis estruturou medidas para intervir nos casos de Febre Puerperal. Em Maio de 1847, seguindo seu método e hipótese, Semmelweis e seus colegas instituíram um programa onde todos que iriam examinar e conduzir mulheres em situação de parto deveriam lavar as mãos com uma solução aquosa de cal clorada. Em Dezembro do mesmo ano, Semmelweis e seus colegas observaram que a taxa de mortalidade por Febre Puerperal caíra de 18% em Abril para 2,45% nos meses seguintes (Gráfico 1). Tal foi o sucesso deste estudo que os órgãos governamentais implementaram programas para que médicos e parteiras a

fim de obrigar estes profissionais a lavarem as mãos antes de conduzir atividades com mulheres em situação de parto. Em contraponto, tais medidas duraram até o fim do mandato de seu cargo, onde tal situação originou um novo aumento dos casos de Febre Puerperal. Alguns anos depois, por motivações associadas à política e de senso comum da sociedade da época que iam contra as hipóteses de Semmelweis, o mesmo acabou sendo enviado a uma instituição para tratamento de pessoas com problemas mentais, onde acabou falecendo em 13 de Agosto de 1865.^{188,190}

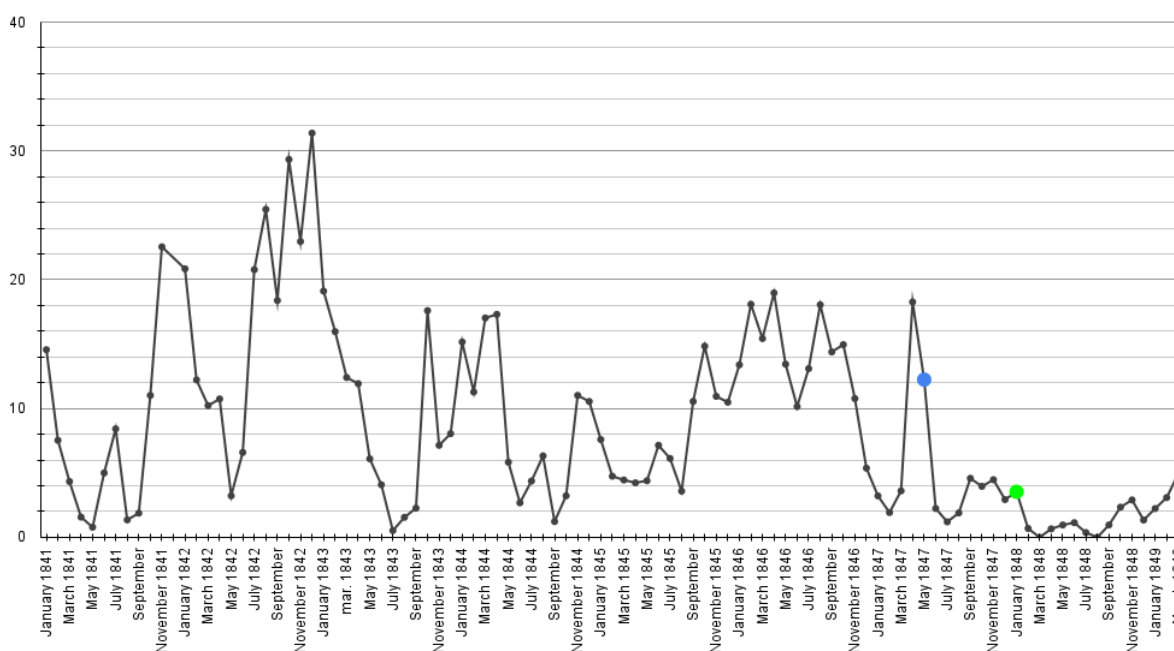


Tabela 1. Taxas de mortalidade por Febre Puerperal do período de Janeiro de 1841 até Março de 1849.

Fonte: O autor, 2024.

O método de Semmelweis se baseou no fato de que, na clínica onde seus colegas médicos efetuaram partos e estudos com cadáveres, a taxa de mortalidade por Febre Puerperal era maior do que na clínica composta apenas por parteiras. De tal forma, Semmelweis elucidou a possibilidade de existir algum “agente venenoso” (como definiu na época) como causador da Febre Puerperal e que, quando os médicos tocavam nos cadáveres, se contaminavam com este tal agente. Desta forma, ele elucidou que a lavagem de mãos poderia eliminar este “agente venenoso” das mãos dos médicos, permitindo que conduzissem partos de forma segura.

Semmelweis também observou a presença de ferimentos purulentos nos pacientes com Febre Puerperal, assim também deduziu que este “agente venenoso” precisava de uma entrada no corpo do paciente para causar a doença.^{187,190,191}

5.2. Criando o DAG

Com base nas suposições e sucesso de suas intervenções, o legado de Ignaz Semmelweis trouxe uma nova ótica para as práticas clínicas da época e enunciou a necessidade de uma medida preventiva que, nos dias atuais, é indispensável. Sua hipótese como causador da Febre Puerperal foi confirmada nos anos seguintes, vindo a elucidar que o “agente venenoso” era, na verdade, um microorganismo ou “agente patogênico”. Tal bactéria, denominada de estreptococos, quando em contato com uma ferida exposta pode entrar na corrente sanguínea do hospedeiro e causar uma infecção generalizada (sepse) de sintomas complexos, podendo levar o paciente a óbito.^{187,192}

Na Figura 23 é possível observar um DAG contendo as prováveis hipóteses que Semmelweis fez em seu raciocínio sobre a causa de morte por Febre Puerperal. Ao analisar as diferenças entre as duas clínicas e observar as práticas de cada uma, Semmelweis deduziu, através do seu método, a provável contaminação das mãos dos médicos através do contato com cadáveres. Tal contaminação de um agente patogênico, quando em contato com uma ferida do puerpério, evolui para um quadro de infecção que pode ser observado a partir de uma ferida contaminada, ou ferida séptica, eventualmente com presença secreção purulenta. Estes sintomas graves observados foram associados e denominados como a Febre Puerperal, que é uma infecção generalizada e possui um alto índice de mortalidade.^{188,192,193}

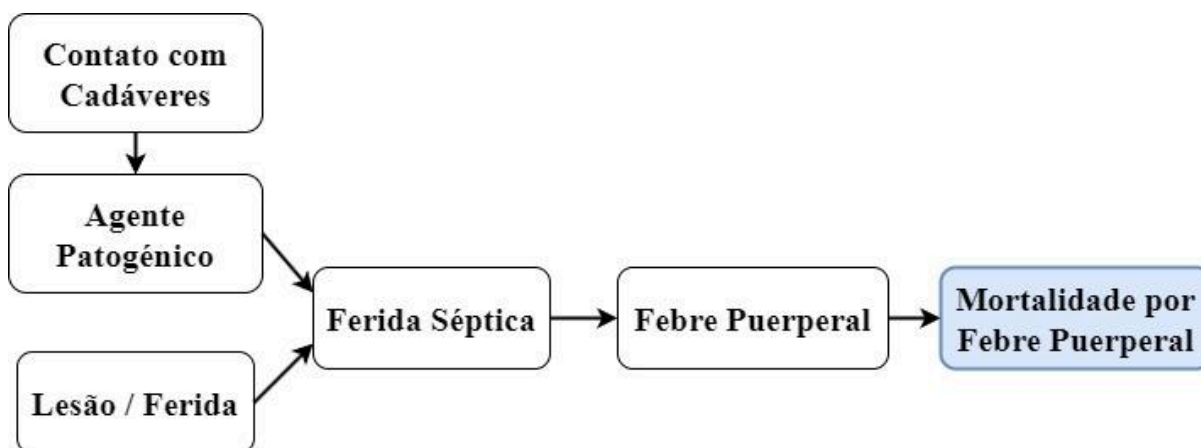


Figura 23. DAG contendo as potenciais hipóteses que Semmelweis fez em seu raciocínio sobre a causalidade da Febre Puerperal.

Fonte: O autor, 2024.

Estudos da época demonstravam o potencial de antissépticos (feitos de cloro) como produtos capazes de higienizar superfícies e, desta forma, eliminar agentes e micróbios indesejados. A partir deste conhecimento, Semmelweis sugeriu que uma intervenção por “higienização das mãos”, com o uso de soluções antissépticas, poderia evitar a Febre Puerperal, proporcionando um prognóstico mais favorável para a sobrevivência de mulheres em situação de pós-parto. De tal forma, a Figura 24 ilustra o raciocínio causal de Semmelweis em intervir com a lavagem de mãos para poder evitar a morte por Febre Puerperal.^{191,194}

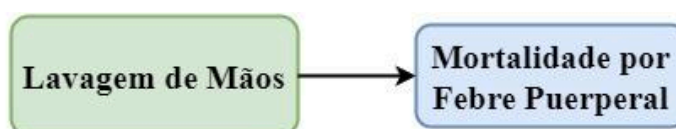


Figura 24. Intervenção através de “Lavagem de mãos” para redução dos casos de Febre Puerperal.

Fonte: O autor, 2024.

O gráfico causal na Figura 25 demonstra um DAG que descreve as relações de causalidade entre os eventos partindo da intervenção “Lavagem de Mãos” e o desfecho de “Mortalidade por Febre Puerperal”. O conjunto de variáveis “Agente Patogénico”, “Ferida do puerpério”, “Ferida Séptica” e “Febre Puerperal” representam nodos que se relacionam de tal forma a funcionarem como variáveis endógenas, *mediators* e *colliders*. O nodo de “Fatores não previsíveis” indica um

conjunto de variáveis exógenas não observadas, de difícil mensuração, mas que corroboram o surgimento de uma lesão no puerpério. O nodo externo “Contato com cadáveres” indica o contato prévio com cadáveres. O nodo intermediário de “Ferida no puerpério” indica o acontecimento de uma lesão durante o processo de parto. O nodo de “Agente Patogênico” indica os níveis de contaminação das mãos por microorganismos. O nodo de “Ferida Séptica” denota a presença de uma ferida contaminada por um agente infeccioso. O nodo de “Febre Puerperal” traz o cenário onde o paciente apresenta níveis de sintomas de Febre Puerperal. O nodo de “Mortalidade por Febre Puerperal” representa o desfecho de interesse, que é: óbito provocado pela doença da “Febre Puerperal”.^{6,192}

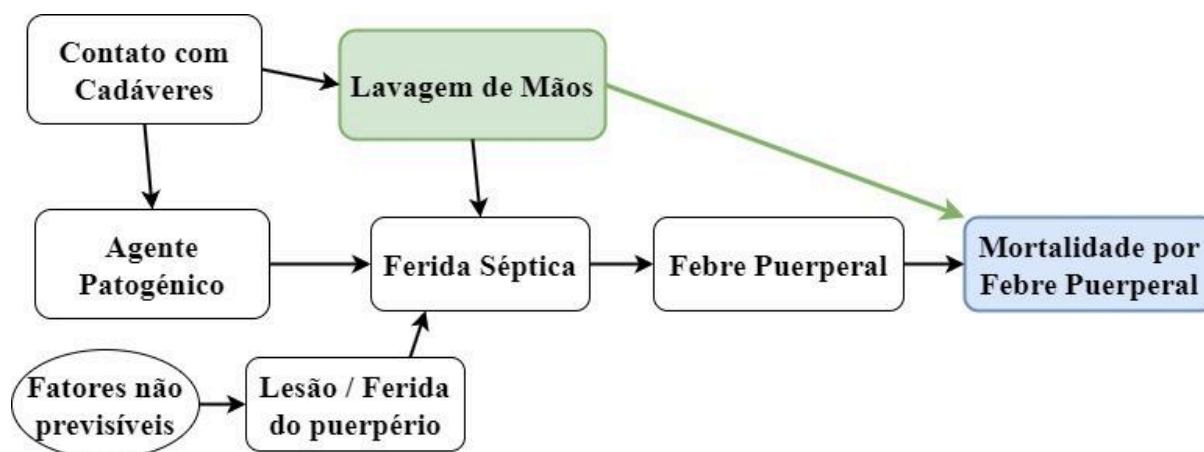


Figura 25. Relações entre nodos denotando a hipótese de causalidade entre “Lavagem de mãos” e o desfecho de “Mortalidade por Febre Puerperal”.

Fonte: O autor, 2024.

5.3. Relações entre nodos e ajustes de caminhos.

No DAG ilustrado pela Figura 25, é possível observar as relações entre os nodos e estabelecer algumas interpretações. As associações entre cada nodo pai e filho é representada por arestas com flechas direcionais e o gráfico causal é acíclico. Desta forma, é possível observar que o DAG possui um caminho de *backdoor*: “Lavagem de Mãos” ← “Contato com Cadáveres” → “Agente Patogênico” → “Ferida Séptica” → “Febre Puerperal” → “Mortalidade por Febre Puerperal”. Tal caminho está bloqueado pelo nodo de “Ferida Séptica”. Desta forma, o critério de *backdoor*

ainda não é satisfeito pela regra. Porém, pelo fato de que o nodo de “Ferida Séptica” também bloquear o caminho causal da intervenção ao desfecho (“Lavagem de Mãos” → “Ferida Séptica” → “Febre Puerperal” → “Mortalidade por Febre Puerperal”), o nodo de “Ferida Séptica”, ao ser desbloqueado, abrirá o caminho de *backdoor* em questão e o critério de *backdoor* se tornará satisfeito.¹²¹

Então, para que a relação de efeito total de causalidade possa ser estabelecida, o caminho de *backdoor* precisa ser tratado de forma a eliminar todas as flechas que chegam na intervenção “Lavagem de Mãos”, o que bloqueia o caminho espúrio de “Lavagem de Mãos” até o desfecho “Mortalidade por Febre Puerperal”. Assim, é possível observar que o nodo de “Contato com Cadáveres” atua como um *confounder* e precisa ser ajustado para satisfazer a regra e bloquear o caminho de *backdoor*. A Figura 26 denota esta nova relação onde o nodo “Contato com Cadáveres” é ajustado e, desta forma, sua relação com “Lavagem de Mãos” é ignorada no DAG.^{106,121}

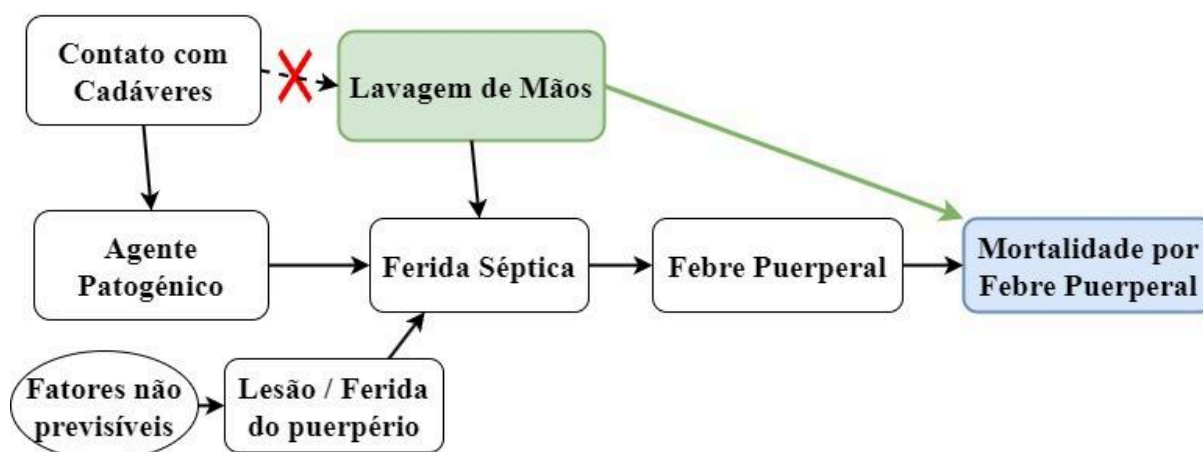


Figura 26. Tratamento do caminho de *backdoor* provocado pela relação entre “Contato com cadáver” e “Lavagem de mãos”.

Fonte: O autor, 2024.

Após o caminho de *backdoor* ser ajustado, este ajuste é minimamente suficiente para permitir que o efeito causal total da intervenção possa ser calculado. Porém, a relação de efeito direto ainda não é possível. O nodo “Ferida Séptica” representa um *collider* no gráfico causal (Figura 26), o que bloqueia e gera *confounding* na inferência de efeito direto da intervenção “Lavagem de Mãos” no desfecho “Mortalidade por Febre Puerperal”. Desta forma, o nodo de “Ferida

Séptica” precisa ser ajustado para que o caminho seja aberto. Se o nodo de “Ferida Séptica” for condicional, o caminho entre “Lavagem de Mãos” e “Mortalidade por Febre Puerperal” será aberto, permitindo a relação de causalidade (Figura 27).^{123,150}

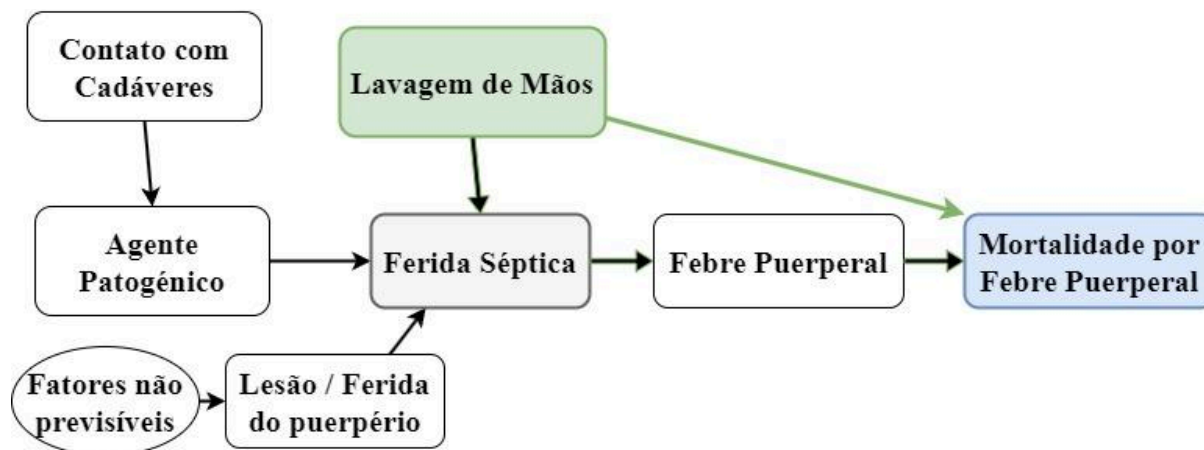


Figura 27. Tratamento da variável “Ferida séptica”.
Fonte: O autor, 2024.

5.4. Calculando efeitos de causalidade

A partir dos ajustes feitos, a fórmula de probabilidade da intervenção, dada pelo *do-operator* através dos critérios de *backdoor* e *frontdoor*, pode ser escrita de forma que a probabilidade da intervenção “Lavagem de Mãos” (“LM”) no desfecho de “Mortalidade por Febre Puerperal” (“MFP”) é igual a probabilidade condicional de “Lavagem de Mãos” em “Mortalidade por Febre Puerperal”, assumindo um sistema linear: $P(\text{“MFP”} | do(\text{“LM”})) = P(\text{“MFP”} | \text{“LM”})$.^{161,165}

O efeito total de causalidade difere do efeito direto. O ajuste no nodo de “Contato com cadáveres” (Figura 26) é suficiente para o cálculo de efeito total pois fecha todos os caminhos enviesados, deixando todos os caminhos causais abertos. Inserindo a variável de “Contato com cadáveres” para regressão junto com a intervenção, deve retornar um coeficiente de regressão r_x associado à exposição “Lavagem de Mãos” que representa o efeito total de todos os caminhos causais.^{6,162}

O efeito direto de causalidade pode ser inferido na Figura 27 visto a variável condicional “Ferida Séptica” dentro do caminho “Lavagem de Mãos” → “Ferida

Séptica → *Febre Puerperal* → *Mortalidade por Febre Puerperal*. Desta forma, adicionando “Ferida Séptica”, “Contato com cadáveres” e a intervenção para regressão, o coeficiente de regressão r_x associado à exposição “Lavagem de Mãos” representará a relação de efeito direto com o desfecho de “Mortalidade por Febre Puerperal”.^{121,164}

No caso de Semmelweis, a intervenção proposta por ele em prática na época de 1847, observou que a Mortalidade por Febre Puerperal reduziu de aproximadamente 18% para 2,45% em um período de 4 meses após a implementação do programa de Higiene de Mãos em todas as clínicas obstétricas da cidade. Considerando que a probabilidade de “Morte por Febre Puerperal” é de 18%: $P("MFP") = 18\%$; é possível deduzir que a probabilidade de “Morte por Febre Puerperal” dado a intervenção por “Lavagem de Mãos” é de 2,45%: $P("MFP" | do("LM")) = 2,45\%$.^{121,163}

5.5. Analisando “Counterfactuals” e contextos atuais

As hipóteses de Semmelweis seguiram um raciocínio lógico e básico. Dentre os questionamentos realizados, certamente buscou identificar fatores que pudessem dar razão ao surgimento das infecções de “Febre Puerperal”. Seguindo uma linha filosófica, o questionamento de Semmelweis provavelmente foi: “O que difere o tratamento das mulheres em situação de parto entre as clínicas para que uma tenha uma taxa de óbitos muito maior que a outra?”. Tal pensamento o permitiu elucidar possibilidades de intervenção e, até mesmo, ter descobertas de sucesso.^{105,158,190}

Este raciocínio filosófico é o cerne para o entendimento da contrafactualidade. Questionamentos como “E se o paciente não tiver uma ferida no puerpério? Ainda terá Febre Puerperal?”, podem parecer questionamentos óbvios e serem respondidos com “Não, o paciente não terá a Febre Puerperal pois o germe precisa de uma abertura para se instalar no hospedeiro e causar a infecção generalizada”. Apesar de tal pensamento atualmente ser intrínseco ao senso comum entre os clínicos e pesquisadores, visto a quantidade de pesquisas posteriores realizadas

sobre o assunto, na época de Semmelweis acreditava-se que a causa era por um envenenamento proveniente do contato com cadáveres, afinal não se tinha conhecimento da existência de microorganismos, que em contato com mulheres em situação de parto geravam a Febre Puerperal.^{47,190,191}

O pensamento comum influencia na forma como pesquisadores criam hipóteses de estudo e, a partir de métodos estatísticos, realizam a afirmação ou não das suas suposições.¹⁹⁵⁻¹⁹⁷ No caso de Semmelweis, até mesmo a pergunta “E se quem for conduzir o parto não lavar as mãos, o paciente terá óbito por Febre Puerperal?” pode ser respondida, em um contexto moderno, com “não necessariamente” e com o complemento de uma série de outras possibilidades.¹⁸³ Este poder que os métodos de inferência causal dão aos pesquisadores de efetuarem questionamentos que não aconteceram é de suma importância para ter-se um entendimento completo do comportamento das intervenções nos desfechos. Até mesmo dá a possibilidade de explorar outras intervenções para raciocinar sobre as relações de causalidade entre fatores e eventos.^{47,105,147,158,198,199}

No contexto atual, a análise de fatores associados às ISCs faz o uso de raciocínios sofisticados, com fundamentos estatísticos modernos para associar as correlações aos desfechos. Índices de risco foram estabelecidos como método para potencial associação de causalidade para ISCs. A partir de fatores de risco, estes critérios geram um *score* que representa a probabilidade de um paciente contrair uma infecção através de uma ferida cirúrgica.¹⁸³⁻¹⁸⁵

Suponha um DAG gerado a partir das características que envolvem estes índices e *scores* para designar uma relação de causalidade de “Antissepsia Cirúrgica” com o desfecho de “Óbito por ISC” (Figura 28). Considerando a intervenção de “Antissepsia Cirúrgica” com o desfecho, algumas características se comportam de tal forma a se relacionarem diretamente com um único evento em comum, o nodo de “ISC”. Essas características representam: a potencial contaminação dos instrumentos, das vestimentas dos cirurgiões ou da própria flora do paciente por um microorganismo; tempo para realização do procedimento; comorbidades do paciente, onde pode ser considerado um conjunto de variáveis como obesidade, diabetes, imunossupressão, má nutrição e uso de cigarro; idade

avançada; histórico de infecções prévias antes do paciente realizar o procedimento; temperatura corporal do paciente durante o procedimento; níveis de oxigenação do sangue; o uso de antibiótico até uma hora antes do procedimento; e a própria ferida cirúrgica (se durante o procedimento foram realizadas as melhores práticas e se foi fechada adequadamente ferida, sem “aberturas” pós cirurgia). O nodo de “ISC” pode evoluir para “Alta melhorada” ou um quadro de infecção generalizada associada ao sítio cirúrgico. A “Infecção Generalizada” pode levar o paciente a “Óbito por ISC”.^{184,185,200–204}

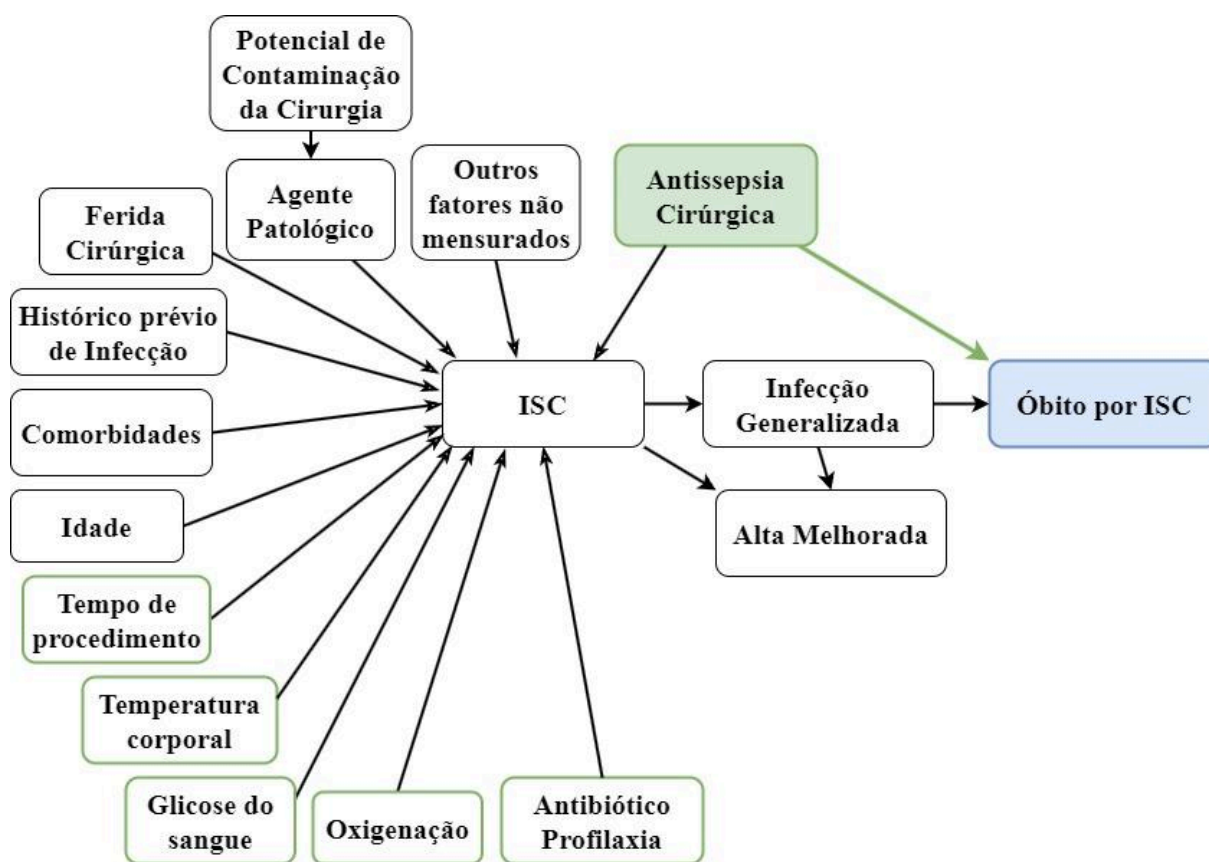


Figura 28. DAG demonstrando a potencial relação de causalidade da variável de “Antissepsia cirúrgica” com relação ao desfecho de “Óbito por ISC”.

Fonte: O autor, 2024.

O trabalho de Semmelweis foi fundamental para que o panorama das práticas clínicas e cirúrgicas fossem aperfeiçoadas. Os esforços pioneiros permitiram com que outros estudos pudessem analisar outras hipóteses para a ocorrência das infecções, reduzindo taxas de infecção e, conseqüentemente, de óbito causado por infecções associadas a procedimentos cirúrgicos. O programa de higienização das

mãos proposto por Semmelweis foi capaz de reduzir drasticamente a taxa de óbitos por Febre Puerperal (entre 5% a 2,5%). Com a aplicação de outras práticas e medidas, a taxa de mortalidade em ISCs atualmente reside em um valor muito menor (inferior a 0,5%). Isso denota a eficácia dos estudos em sugerir boas práticas clínicas e a adoção de medidas protetivas e preventivas para cuidados com os pacientes. Tais práticas são registradas em protocolos, normativas, índices e *scores* de risco. Funcionam de tal forma a fornecer uma metodologia padrão e universal para mensurar os riscos de um procedimento cirúrgico.^{188,191,192,202}

Os modelos e hipóteses causais podem ser testados a partir da criação de modelos próprios ou de pacotes já existentes em linguagens de programação modernas. O pacote disponível para a linguagem de programação *R*, denominado “*dagitty*”, permite aos usuários analisar e manipular DAGs, sendo este o pacote criado pelos alunos de Judea Pearl.²⁰⁵ Outros pacotes existentes para a linguagem de programação *R* como por exemplo “*CausalQueries*” permitem aos usuários testarem vários outros efeitos de inferência causal, além da metodologia explanada.²⁰⁶

6. DISCUSSÃO

Desde registros de filósofos antigos até os tempos modernos, civilizações buscaram entendimento do que seriam relações de causa e efeito. Pensamento intrínseco aos seres humanos, a busca pela compreensão de conceitos de causalidade foi constante em diversos registros da humanidade e variou conforme o passar das eras.^{1,2,207,208}

Em tempos recentes, o ECR foi fundamentalmente considerado como principal método para estabelecer relações de causalidade. Porém, tal método pode acabar não sendo viável ou ético, pela possibilidade de expor pacientes a riscos ou pelo fato de não ser possível intervir em variáveis como câncer, obesidade, idade etc. Um ECR também pode acabar sendo custoso ou de alta complexidade, sendo este último um fator crítico para a possibilidade de surgimento de vieses de estudo.^{47,209} Mesmo assim, estudos epidemiológicos se restringiram unicamente a ECRs para denotar relações de causalidade.²¹⁰

Estudos observacionais surgem como alternativa importante para casos em que o ECR não é viável, porém as abordagens convencionais utilizadas em estudos epidemiológicos através de tais dados observacionais são incapazes de detectar e demonstrar a transmissão de efeitos de causalidade.^{115,209} Desde os critérios enunciados por Bradford Hill até estudos que analisam os fatores de risco associados às doenças elucidam apenas informações meramente correlativas. Estas covariáveis podem conter relações de efeito causal enviesadas, sendo assim, incapazes de esclarecer as relações de efeito e causa.^{115,211}

Com o intuito de viabilizar relações efetivamente causais em estudos observacionais, os métodos de inferência causal surgem como uma ferramenta capaz de ilustrar os fluxos de causalidade através de um DAG. Considerando que tais métodos são de custo reduzido se comparados à realização de ECRs, as técnicas de inferência causal podem também elucidar diferenças entre as relações espúrias e causais. O método do *do-calculus* ajusta o gráfico causal e calcula os efeitos de causalidade, contendo as mesmas qualidades da randomização em ECs, só que em dados observacionais.⁴⁷

Aspecto interessante a ser mencionado é que a metodologia não necessita “conhecer” ou representar todas as variáveis presentes em um SCM para inferir alguma relação de efeito causal. Com a representação visual em um DAG é possível tratar os caminhos espúrios (*backdoor*) e, assim, extrair informações importantes para o entendimento de causalidade das hipóteses de estudo.^{47,121}

Quando este conhecimento é aplicado no contexto de ECRs, os métodos podem complementar e validar o conhecimento adquirido pelos ensaios. A exibição das hipóteses de estudo de um ECR em um DAG pode apoiar e qualificar as relações entre as variáveis, elucidando as correlações e facilitando o entendimento de efeito causal entre intervenção e desfecho.^{107,212-215} Também, métodos de inferência causal podem vir a reforçar hipóteses que viabilizem ECRs, elucidando a necessidade ou possibilidade de mais estudos antes da realização de tais ensaios.²¹⁶ Por fim, estudos de inferência causal são comumente rápidos e possuem custos ínfimos se comparados aos custos de um ECR.^{47,106,216}

Os métodos de inferência causal, por serem realizados em dados observacionais, possuem limitações que estão diretamente associadas a questões amplamente conhecidas de estudos observacionais, que podem impedir com que a meta análise seja feita. Tais limitações provocam vieses, impedindo que a compreensão da transmissão de causalidade possa ser inferida. Os erros presentes nos dados podem enunciar questionamentos tanto quanto à qualidade quanto à integridade dos mesmos. Também pode ser considerada a capacidade do conjunto de informações utilizadas no estudo de representar um escopo generalizado, elucidando a necessidade da realização de ECRs para viabilizar as relações de causa e efeito ou, quando possível, a aquisição de mais dados observacionais.^{106,108,148,152}

A tipagem dos dados pode afetar diretamente como os resultados são interpretados e pode agregar em complexidades quando se deseja validar a generalização do modelo. Caso um modelo tenha sido concebido com uma variável contínua para sua representação, por exemplo, e, em situação de validação externa, caso esta variável seja categórica ou binária, o modelo pode ser ajustado para viabilizar os fluxos de causalidade, porém é importante notar que isso implicará em

mudanças importantes para validação e interpretação da generalização do modelo. O uso de variáveis contínuas permite melhor interpretação e viabilização de técnicas de regressão. Conduz pesquisadores a mensurar as relações causais com maior assertividade e qualidade. Em caso de impossibilidade de se usar variáveis contínuas, os DAGs viabilizam que qualquer tipagem de dados possa ser utilizada a fim de se avaliar a inferência de fluxos de causalidade.⁴⁷

A metodologia ainda pode imitar o desenho de um ECR envolvendo um ensaio hipotético, denominado como *Target Trial Emulation*. Esta abordagem envolve a especificação de um ensaio “alvo”, incluindo os seus critérios de elegibilidade, estratégias de tratamento, procedimentos de atribuição e resultados, e depois emular este ensaio utilizando dados do mundo real. Ao fazê-lo, gera um fortalecimento das inferências causais extraídas de dados observacionais. Este método enfatiza a importância de definir claramente a questão causal e abordar sistematicamente as fontes de preconceito e confusão, como se faria no planejamento de um ECR.²¹⁷

Desta forma, estudos epidemiológicos associados ao combate das IRAS podem: viabilizar novas hipóteses e intervenções de estudo baseados em diversos tipos de evidências, mesmo quando não adequadas a um ECR; elucidar novas estratégias para avaliar relações de causa e efeito, descartando variáveis de estudo que transmitam correlações espúrias; incluir desenhos de estudo que não correspondem necessariamente ao objetivo da tese, isto é, hipóteses contrafactuais; e por fim, proporcionar uma reflexão crítica sobre quais os fatores que efetivamente tem transmissão de efeito causal, mesmo que isso vá contra questões culturais ou de preconceito prévio.^{210,218} Esta tese abordou metodologias de inferência causal que visam esclarecer eventuais dúvidas sobre processos, elucidando como que o *do-operator* pode, a partir de uma hipótese e um DAG, indicar relações de efeito causal de uma intervenção em um desfecho.

7. RELATÓRIO SINTÉTICO DO PRODUTO 1 DA DISSERTAÇÃO

7.1 – Título

FRAMEWORK PARA ESTUDOS DE INFERÊNCIA CAUSAL EM ESTUDOS OBSERVACIONAIS SOBRE INFECÇÕES RELACIONADAS À ASSISTÊNCIA À SAÚDE.

7.2- Descrição

O presente produto é um *framework* que deve servir como ferramenta para orientar pesquisadores a realizarem estudos de inferência causal na área das IRAS de forma adequada e padronizada. Contendo diretrizes e um exemplo, esta ferramenta também deve esclarecer eventuais dúvidas e servir como referencial base para que novos estudos de inferência causal ocorram na área das infecções hospitalares.

7.3 - Aplicabilidade do produto

A aplicabilidade do produto pode ser realizada de forma imediata. O conhecimento registrado neste framework pode ser aplicado em qualquer instituição, desde que associado às IRAS.

7.4 - Inserção social

Este produto deve contribuir significativamente para que novos estudos de inferência causal na área das IRAS possam ocorrer. Pesquisadores podem usar este framework como guia base para estudar relações de causa e efeito de interesse na área das IRAS e, assim, podem elucidar novas descobertas ou reafirmar conhecimentos previamente descobertos.

7.5 – Previsão de publicação [onde e como será publicado e divulgado]

A tese será publicada na biblioteca da UFRGS e será divulgada a partir de futuros estudos na área das IRAS.

8. RELATÓRIO SINTÉTICO DO PRODUTO 2 DA DISSERTAÇÃO

8.1 – Título

How Causal Inference Methods Could Have Aided Ignaz Semmelweis in the Puerperal Fever Case

8.2- Descrição

O produto é um artigo científico que busca esclarecer alguns aspectos de inferência causal e, a partir do caso de Febre Puerperal estudado por Ignaz Semmelweis no século, elabora uma narrativa argumentativa de como Semmelweis poderia ter se beneficiado de tais métodos no intuito de conseguir persuadir a comunidade médica de sua época. Por fim, conclui-se que novos estudos de inferência causal também podem usar de tais artifícios para elucidar ou solidificar o conhecimento sobre relações de causalidade em IRAS.

8.3 - Aplicabilidade do produto

O produto tem aplicabilidade imediata após a publicação e deve proporcionar um entendimento fundamental para o acontecimento de estudos de inferência causal na área das IRAS.

8.4 - Inserção social

O produto deve esclarecer o entendimento que pesquisadores têm sobre causalidade nas IRAS, propondo que comunidades acadêmicas busquem melhor entendimento sobre tais técnicas.

8.5 – Previsão de publicação [onde e como será publicado e divulgado]

O artigo será publicado no American Journal of Infection Control e será divulgado pela própria revista. A previsão de publicação é em 05/06/2024.

9. ARTIGO CIENTÍFICO: “How Causal Inference Methods Could Have Aided Ignaz Semmelweis in the Puerperal Fever Case”

Introduction

In the 19th century, puerperal fever, also known as childbed fever, was a major cause of maternal death. Ignaz Semmelweis, a Hungarian physician, observing the significantly higher maternal mortality rates in the clinic staffed by doctors compared to the midwives' clinic, hypothesized that the puerperal fever was transmitted via cadaveric particles from the hands of medical students from autopsies performed. He hypothesized that handwashing with chlorinated lime solutions could drastically reduce mortality rates. However, despite his success, the lack of a theoretical framework and empirical methods required for establishing causal relationships hindered widespread acceptance of his findings.¹

According to the American Centers for Disease and Control (CDC) 1 in every 31 inpatients are at risk of healthcare-associated infection (HAI).² This risk is universal, and the burden of these infections are even higher in low and middle-income countries.³ HAIs are preventable, and risk factors related to infection may be endogenous or exogenous or, most often a combination of both. Causality link to HAI is crucial for prevention. However there is scarce literature of causality and infection control.^{4,5}

Most epidemiological observational studies focus on establishing associations to enlight elements of causality. Causation presupposes that the exposure produces the effect.⁶ Traditionally, causal associations and effects are most accurately taken from randomized controlled trials (RCTs) where common confounders can be controlled, however, this approach can not be used in all cases, and sometimes it is not possible in the field of infection prevention.⁷

Modern causality benefits from tools that can be used in observational studies to

simulate the effects of randomization, controlling for confounders and denoting cause-and-effect relationships.^{8,9} In the context of HAIs, the opportunities for conducting causal inference studies are vast.¹⁰

Puerperal sepsis case

In 1847, Ignaz Semmelweis investigated the causes of mortality in postpartum childbirth. Based on methodological reasoning, he hypothesized that handwashing with chlorinated lime solution among healthcare clinicians and midwives could significantly reduce the incidence of puerperal fever.¹

Despite substantial evidence supporting his hypothesis, due to the lack of knowledge about pathogenic microorganisms and the absence of statistical and methodological tools to demonstrate causal relationships, Semmelweis faced resistance from the medical community about his findings.¹¹ The use of a causal diagram could have held him a greater advantage in showing that his hypotheses were somehow causal, and not merely correlative, even if we consider biases and lack of modern knowledge to it.^{9,12}

Causal Inference

The methodology is based on a concept where the causal assumptions are represented in a simple format. Such assumptions are displayed in a Directed Acyclic Graph (DAG) that qualifies the causal theories of interest, elucidating confounders and enabling the inference of causality transmission between variables.¹³

The *do-calculus* provides a theoretical framework for drawing causal inferences from observational data.¹⁴ It consists of a set of rules for adjusting spurious paths, manipulating conditional probabilities for hypothetical interventions and handling confounders.¹⁵ It allows

researchers to estimate causal effects even when RCTs are not feasible.⁹

The causal graph should be constructed from hypotheses previously reasoned based on existing knowledge in the literature, even if they represent simplistic assumptions of a much more complex context. By exhibiting relationships between variables as nodes with directed edges, each directed edge represents the potential causal relationship between one variable to another and explicit causal hypotheses of interest.^{9,16}

In most cases, the generated DAG may not be correct and it may contain biases related to data that are hard to detect. However, the explicit visualization of the hypotheses in a DAG allows the researcher to extract important information such as identifying which paths need to be adjusted, which variables are merely correlational, and thus, which variables are appropriate for regression, reducing the possibility of study confounding.¹⁵

The process of creating and adjusting a DAG can become considerably complex. The proper selection of variables that should be included in the model is a delicate task. Researchers commonly consider a wide selection of variables, which can increase biases and create spurious paths that need to be adjusted.¹⁷

The quality of inferences drawn from observational causal thesis depends on various factors highly reliant on the quality and integrity of the data. The researcher's knowledge, beliefs and ability to measure errors can lead to bias that are hard to be measured. Nevertheless, strong study hypotheses, when combined with extensive knowledge about a specific subject, can generate relevant causal diagrams for understanding the cause-and-effect relationships that an intervention has on an outcome.^{4,9,18}

Application to Semmelweis's Case

A DAG representing Semmelweis's hypothesis would have included nodes following

the standard beliefs at the time. The nodes in the graph shown on Figure 1 could be the practice of handwashing as the intervention (“Handwashing”), cadaveric particles on the clinician hand (“Cadaveric Particles”), the presence of a puerperal wound due to birth procedure (“Puerperal Wound”), a contaminated wound with the presence of infectious matter (“Septic Wound”), puerperal fever symptoms (“Puerperal Fever”), and mortality by puerperal fever (“Death by Puerperal Fever”). The directed edges coming into “Septic Wound” from the nodes “Handwashing”, “Cadaveric Particles” and “Puerperal Wound” represent the causal assumptions that Semmelweis’s did as he observed that all inpatients with puerperal fever had a septic wound. Therefore, with the presence of a septic wound, it evolves into a puerperal fever condition (“Puerperal Fever”), that could lead the inpatient to death (“Death by Puerperal Fever”). Finally, the direct edge from “Handwashing” to “Death by Puerperal Fever” shows his intentions of intervening with handwashing on the outcome of mortality by puerperal fever.^{1,15}

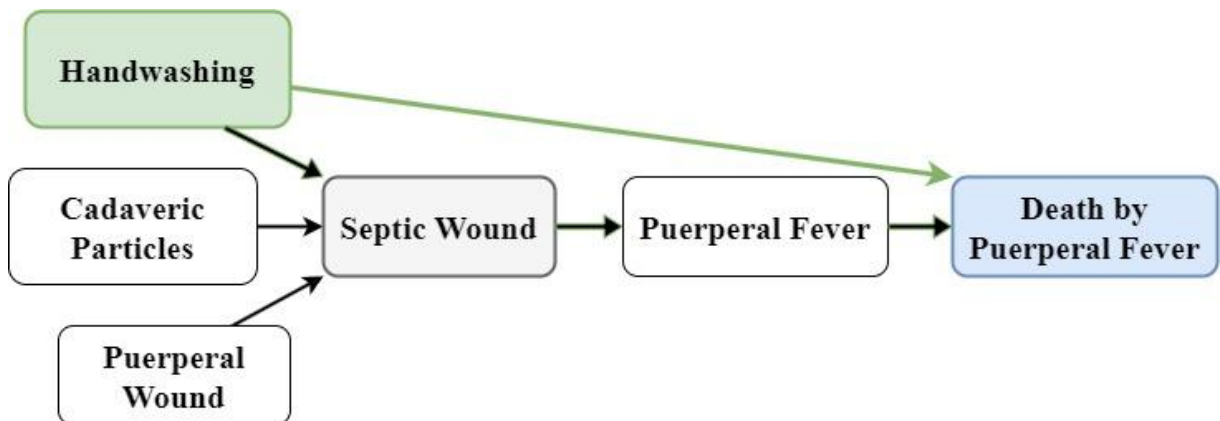


Figure 1. A DAG exhibiting the possible assumptions that Semmelweis made on his hypotheses for “handwashing reduces mortality caused by Puerperal Fever”.

Identifying Confounding Factors

Using DAGs as a simple tool in order to express and explicit his assumptions, Semmelweis could have persuaded the medical community at his time. By identifying potential confounders and adjusting for the “Septic Wound” node, Semmelweis could have shown to the medical community the causal effect that intervention (“Handwashing”) has on the outcome (“Death by Puerperal Fever”), which could have strengthened his causal claims.¹⁴

Estimating Causal Effects Using *do-calculus*

Semmelweis could have formally estimated the causal effect of the intervention “Handwashing” (“HW”) on the incidence of “Death by Puerperal Fever” (“DPF”) by using the *do-operator*, $do(“HW”)$. He was able to reduce the rates of “Death by Puerperal Fever” from 18% to 2.45% in the subsequent months.¹⁹ Assuming a linear distribution of the variables, it would be safe to assume that the probability of the intervention would be the same as the conditional probability: $P(“DPF” | do(“HW”)) = P(“DPF” | “HW”)$.¹⁵ Even considering the absence of a RCT, the DAG and the *do-calculus* method should provide enough qualitative and quantitative evidence to support his hypothesis.⁹

Discussion

The benefits of a causal inference are promising. It gives a deeper understanding of a specific subject and, even if the causal model is wrong, it is still somehow useful. The ability of answering interventional queries are inherited from a standard human perspective and modern causal inference methods are capable of simulating these scenarios using observational data.⁹

The *do-calculus* method thus an explicit demonstration of his hypotheses on a DAG

could have provided clear, formal representation for Semmelweis's assumptions. The estimation of causal effect that handwashing has on reducing puerperal fever mortality could be enlightened by these methods, since they allow the proper identification and adjustment of confounders. Furthermore, causal inference methods should have provided a more persuasive argument to Semmelweis's that hand hygiene practices prevent puerperal fever and should have helped him overcome the skepticism of his time.^{1,9}

Artificial intelligence (machine learning) and its complex interactions could be strongly benefited from causal analysis, since machine learning and deep learning models are incapable of answering interventional queries. Still, machine learning based methods could accelerate clinical processes by identifying patterns and predicting outcomes, providing a better understanding of the associations between variables. It could also improve confounder handling and reduce biases in observational studies.^{9,18}

Target trial emulation can be used as a framework for causal inference in observational studies. The methodology aims to mimic the design of a RCT as closely as possible. This approach involves specifying a hypothetical or "target" trial, including its eligibility criteria, treatment strategies, assignment procedures, and outcomes, and then emulating this trial using real-world data. By doing so, researchers attempt to replicate the conditions of an RCT, thereby strengthening causal inferences drawn from observational data. This method emphasizes the importance of clearly defining the causal question and systematically addressing sources of bias and confounding as one would in planning an RCT.²⁰

The *do-calculus* requires a solid understanding of causal theories and methodologies, as well as careful consideration of the assumptions underlying the data and the causal models. These methods highlight the growing sophistication in causal analysis, enabling researchers to draw more reliable causal inferences in situations where traditional experimental approaches

are not feasible. In Semmelweis's theoretical case, *do-calculus* provides a formal mechanism for deriving causal effects from observational data using causal graphs, while alternatives like the target trial emulation would focus on the practical design of observational studies to replicate the conditions of RCTs. *do-calculus* is more abstract and requires a solid understanding of the causal model, making it suitable for theoretical investigations and complex causal questions.¹⁵

A recent systematic review of longitudinal infectious diseases studies found lack of standardization of the studies that implemented modern causal analysis, due to heterogeneity in the methods that were applied. Education on specific analysis tools or implementation guidelines, interdisciplinary collaborations between statisticians and infectious diseases fellows, could break some of the barriers for causal inference implementation.⁴

Causal inference frameworks have been proposed in many fields. In urology, Shapiro et al, introduced causal diagrams and applied them to guide clinical decisions in renal cell carcinoma. In many settings in the urology field it is difficult to make individual decisions based on RCT. The authors proposed using this framework to combine RCT with other study sources to compute new statistical estimates for individual decision making. In a retrospective cohort study, Kroenke et al, used a DAG to evaluate how BMI affects survival in colorectal cancer patients.^{21,22} They addressed issues of reverse causality, collider bias, and used acyclic graphs diagrams to deal with confounding bias. Other study in oncology, underscored the importance of using causal diagrams and integrating diverse data sources to improve the precision and reliability of individualized treatment decisions.²³ In psychology, Rohrer (2018) established a systematic approach to improve causal inference in observational research using graphical causal models aiding researchers to apply this models with better control of confounding and biases.¹²

JAMA's recent article proposes a framework containing a general methodology for causal inference studies in observational studies. This framework recommends answering six core questions that should serve as a guideline for best practices. If possible, these questions must be answered in all observational studies about causal inference methods to better understand the study's objectives and results.²⁴

A causal inference framework in the area of infectious diseases, following multiple suggestions of good practices, could lead researchers to implement processes that lead to greater harmonization between causal inference in observational studies, allowing readers to adequately understand how the analyzes are conducted. Additionally, it could reduce researchers' errors by providing a standard approach to developing causal inference interpretation in observational studies.

Conclusion

If Ignaz Semmelweis relied on modern causal inference methods for observational data, he could have probably persuaded the medical community at his time to accept his discoveries.²⁵ A robust causal framework for validating and communicating findings could potentially lead to early acceptance and implementation of life-saving practices across healthcare communities. The DAG could enlighten relevant causal knowledge and could have given enough tools to Semmelweis to show his hypotheses by exhibiting the causal effects of "Handwashing" on the sepsis known as "Puerperal Fever".

References

1. Tyagi U, Barwal KC. Ignac Semmelweis-Father of Hand Hygiene. *Indian J Surg* [Internet]. 2020 Jun;82(3):276–7. Available from: <http://dx.doi.org/10.1007/s12262-020-02386-6>
2. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated

- infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control* [Internet]. 2008 Jun;36(5):309–32. Available from: <http://dx.doi.org/10.1016/j.ajic.2008.03.002>
3. Ling ML, Apisarnthanarak A, Madriaga G. The Burden of Healthcare-Associated Infections in Southeast Asia: A Systematic Literature Review and Meta-analysis. *Clin Infect Dis* [Internet]. 2015 Jun 1;60(11):1690–9. Available from: <http://dx.doi.org/10.1093/cid/civ095>
 4. Hufstedler H, Rahman S, Danzer AM, Goymann H, de Jong VMT, Campbell H, et al. Systematic Review Reveals Lack of Causal Methodology Applied to Pooled Longitudinal Observational Infectious Disease Studies. *J Clin Epidemiol* [Internet]. 2022 May;145:29–38. Available from: <http://dx.doi.org/10.1016/j.jclinepi.2022.01.008>
 5. Allegranzi B, Bagheri Nejad S, Combescure C, Graafmans W, Attar H, Donaldson L, et al. Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis. *Lancet* [Internet]. 2011 Jan 15;377(9761):228–41. Available from: [http://dx.doi.org/10.1016/S0140-6736\(10\)61458-4](http://dx.doi.org/10.1016/S0140-6736(10)61458-4)
 6. Paynter S. Incorporating Transmission Into Causal Models of Infectious Diseases for Improved Understanding of the Effect and Impact of Risk Factors. *Am J Epidemiol* [Internet]. 2016 Mar 15;183(6):574–82. Available from: <http://dx.doi.org/10.1093/aje/kwv234>
 7. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG* [Internet]. 2018 Dec;125(13):1716. Available from: <http://dx.doi.org/10.1111/1471-0528.15199>
 8. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* [Internet]. 2005;95 Suppl 1:S144–50. Available from: <http://dx.doi.org/10.2105/AJPH.2004.059204>
 9. Pearl J, Mackenzie D. The book of why: the new science of cause and effect [Internet]. Basic books; 2018. Available from: http://repo.darmajaya.ac.id/4847/1/The%20book%20of%20why_%20the%20new%20science%20of%20cause%20and%20effect%20%28%20PDFDrive%20%29.pdf
 10. Hufstedler H, Matthay EC, Rahman S, de Jong VMT, Campbell H, Gustafson P, et al. Current trends in the application of causal inference methods to pooled longitudinal observational infectious disease studies-A protocol for a methodological systematic review. *PLoS One* [Internet]. 2021 Apr 29;16(4):e0250778. Available from: <http://dx.doi.org/10.1371/journal.pone.0250778>
 11. Persson J. Semmelweis's methodology from the modern stand-point: intervention studies and causal ontology. *Stud Hist Philos Biol Biomed Sci* [Internet]. 2009 Sep;40(3):204–9. Available from: <http://dx.doi.org/10.1016/j.shpsc.2009.06.003>
 12. Hernán MA, Robins JM. Causal Inference: What If [Internet]. Chapman & Hall/CRC; 331 p. Available from: <https://play.google.com/store/books/details?id=mOh7zweEACAAJ>
 13. Greenland S, Pearl J. Causal Diagrams. 2007 [cited 2024 Feb 24]; Available from: <https://escholarship.org/uc/item/7tn3p6jx>
 14. Pearl J. Causality: Models, Reasoning and Inference [Internet]. Cambridge University Press; 2009. Available from: <https://play.google.com/store/books/details?id=6blRkgEACAAJ>
 15. Pearl J, Glymour M, Jewell NP. Causal Inference in Statistics: A Primer [Internet]. John Wiley &

- Sons; 2016. 160 p. Available from:
<https://play.google.com/store/books/details?id=I0V2CwAAQBAJ>
16. Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol* [Internet]. 2002 Feb;31(1):163–5. Available from: <http://dx.doi.org/10.1093/ije/31.1.163>
 17. Rohrer JM. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science* [Internet]. 2018 Mar 1;1(1):27–42. Available from: <https://doi.org/10.1177/2515245917745629>
 18. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* [Internet]. 2016 Apr 15;183(8):758–64. Available from: <http://dx.doi.org/10.1093/aje/kwv254>
 19. Semmelweis I (1861) [Internet]. The James Lind Library. 2010 [cited 2024 Feb 25]. Available from: <https://www.jameslindlibrary.org/semmelweis-i-1861/>
 20. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA* [Internet]. 2022 Dec 27;328(24):2446–7. Available from: <http://dx.doi.org/10.1001/jama.2022.21383>
 21. Kroenke CH, Neugebauer R, Meyerhardt J, Prado CM, Weltzien E, Kwan ML, et al. Analysis of Body Mass Index and Mortality in Patients With Colorectal Cancer Using Causal Diagrams. *JAMA Oncol* [Internet]. 2016 Sep 1;2(9):1137–45. Available from: <http://dx.doi.org/10.1001/jamaoncol.2016.0732>
 22. Shapiro DD, Msaouel P. Causal Diagram Techniques for Urologic Oncology Research. *Clin Genitourin Cancer* [Internet]. 2021 Jun;19(3):271.e1–271.e7. Available from: <http://dx.doi.org/10.1016/j.clgc.2020.08.003>
 23. Msaouel P, Lee J, Karam JA, Thall PF. A Causal Framework for Making Individualized Treatment Decisions in Oncology. *Cancers* [Internet]. 2022 Aug 14;14(16). Available from: <http://dx.doi.org/10.3390/cancers14163923>
 24. Dahabreh IJ, Bibbins-Domingo K. Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals. *JAMA* [Internet]. 2024 May 9; Available from: <http://dx.doi.org/10.1001/jama.2024.7741>
 25. Gillies D. Hempelian and Kuhnian approaches in the philosophy of medicine: the Semmelweis case. *Stud Hist Philos Biol Biomed Sci* [Internet]. 2005 Mar;36(1):159–81. Available from: <http://dx.doi.org/10.1016/j.shpsc.2004.12.003>

10. CONCLUSÃO

Causalidade em epidemiologia e estudos clínicos são fundamentalmente de origem probabilística. As análises partem de pressupostos relacionados a identificar a probabilidade que uma intervenção tem em alterar a ocorrência de um desfecho ou evento de interesse.²¹⁹

ECRs foram mantidos pela comunidade científica como um padrão ouro para estudos causais, ressaltando-se que, apesar de suas qualidades, suas limitações são amplamente conhecidas. Questões éticas, que podem colocar a vida de pacientes em risco, surgem como fatores que restringem a elaboração de teses causais, não permitindo com que certos elementos de causa e efeito possam ser elucidados a partir de tais intervenções.^{46,100,215}

Partindo deste contexto, estudos observacionais emergem como alternativa importante e, com o uso dos métodos explanados, permitem com que seja possível extrair informações importantes que viabilizem a interpretação de relações causais. Métodos de inferência causal dão aos pesquisadores uma ferramenta poderosa. Permitem com que simulem cenários interventivos em dados observacionais ou chegar a mesmas conclusões que ECRs, só que com relações de custo e tempo mais efetivos.^{47,220,221}

Um gráfico causal deve ser construído a partir de hipóteses raciocinadas previamente, mesmo que representem suposições simplistas de um contexto complexo. Conforme novas características ou eventos forem coletados a partir da leitura de dados, o DAG pode ser refinado a fim de contemplar tais questões. As conexões entre os nodos denotam associações de causa e efeito e os caminhos gerados permitem diferenciar fluxos causais das correlações espúrias.^{124,134,222}

Mesmo permitindo inferir relações entre características, um DAG é incapaz de demonstrar quais os vieses não associados às relações entre as variáveis que existem no modelo. Isto implica que todo gráfico causal gerado está, de alguma forma, incorreto, mas sem ele, todas as suposições realizadas seriam implícitas. Um DAG, além de explicitar as hipóteses, ainda, sim, permite identificar variáveis que

não deveriam estar presentes para regressão, reduzindo vieses e a possibilidade de *confounding*.^{121,219,223}

Em caso de existirem discordâncias na construção de um gráfico causal, que impossibilite a interpretação dos resultados, a análise e construção do DAG pode ser refeita, assim, gerando novos gráficos causais. Isso permite comparações e novas conclusões, elucidando interações entre variáveis e enriquecendo as potenciais relações de causalidade do modelo, por meio de implicações significativas.^{215,224,225}

As metodologias explanadas de caminhos de *backdoor* e *frontdoor* assim como as relações de *d-separation* permitem identificar relações espúrias. *d-separation* auxilia a identificar no modelo um conjunto de relações que precisam ser ajustadas, sem a necessidade de conhecer as funções que regem as relações entre as variáveis. Os critérios de caminhos de *backdoor* e *frontdoor* se beneficiam dos conceitos de *d-separation* e elucidam quais fluxos são causais e quais são os caminhos não causais.^{121,123}

Vieses não associados aos caminhos são complexos de serem tratados e denotam parte importante na análise dos dados. Tais vieses podem implicar em interpretações errôneas e precisam ser tratados sempre que possível. Em casos em que não seja possível mensurar parâmetros, estes devem estar representados no DAG com nodos de seleção (S) para denotar *confounding* nas relações de causalidade (*selection bias*). Aspecto importante também a ser mencionado é que, quando um viés é tratado com métodos de estratificação, isso pode introduzir *selection bias* no modelo causal. Caso nenhuma variável possa ser mensurada ou estimada, o modelo causal não pode ser testado.^{121,150,153}

As análises de efeito causal (direto e total) podem ser realizadas após o DAG ser tratado e ajustado adequadamente. Métodos de regressão auxiliam modelos causais lineares a ter uma equação descritiva que identifica e mensura as relações de efeito causal. As “expectativas” para as probabilidades das relações entre as variáveis de interesse determinam os efeitos direto e total mediante distribuições lineares.^{121,164,165}

A qualidade das inferências depende diretamente da integridade dos dados usados e do conhecimento estatístico de quem faz a análise e estimação dos erros provocados por *confounders* não lineares. A análise dos dados deve considerar a verificação de fatores que resultem em *selection bias* endógeno, o que pode gerar *confounding* com as relações causais de interesse. O ajuste de variáveis só deve ser feito se necessário, nem todas as variáveis devem ser ajustadas se não podem implicar na criação de *confounding*. Caminhos de *backdoor* precisam ser tratados antes da análise e inserção dos dados.^{106,108,131,162}

O processo de criar e ajustar um DAG pode se tornar consideravelmente complexo. A escolha adequada das variáveis que devem estar presentes no modelo é uma tarefa delicada e, sempre que possível, deve ser realizada antes da criação dos grafos causais. Pesquisadores comumente consideram uma seleção ampla de variáveis, o que pode aumentar vieses e criar novos caminhos de *backdoor* que necessitam ser tratados. Após a realização dos ajustes, métodos de regressão devem “filtrar” nodos correlativos, que podem ser desconsiderados no estudo.^{47,131}

Realizar estudos de causalidade é uma tarefa complexa que requer a tomada de muitas decisões multifacetadas de difícil exatidão. A qualidade das inferências extraídas de estudos causais dependem de diversos fatores altamente dependentes da qualidade e integridade dos dados e da organização das hipóteses de quem realiza o estudo, podendo ser enviesada por crenças ou a habilidade do pesquisador em mensurar tais erros. Mesmo assim, hipóteses de estudo fortes, quando combinadas com um amplo conhecimento sobre um determinado assunto, podem originar diagramas causais relevantes para uma compreensão das relações de causa e efeito que uma intervenção tem sobre um desfecho.^{131,199}

Novos estudos podem surgir a partir dos métodos explanados e, possivelmente, validando o *framework* proposto.¹¹⁰ As metodologias de inferência causal são estudadas ainda com certo ceticismo, contudo precisam estar presentes e com grau de domínio metodológico cada vez maior pelas comunidades acadêmicas e entidades de pesquisas clínicas.^{107,119,226} Estudos futuros podem não só estabelecer conclusões causais sofisticadas sobre infecções hospitalares, mas também validar as práticas existentes, levando a melhores prognósticos para os

pacientes e otimizando os processos relacionados à prestação de cuidados de saúde.^{47,110,209}

11. REFERÊNCIAS

1. Kant I. Critique of pure reason. 1781. Modern Classical Philosophers, Cambridge, MA: Houghton Mifflin [Internet]. 1908;370–456. Available from: <https://www.stephenhicks.org/wp-content/uploads/2013/03/kanti-critique-pure-reason-2nd-preface.pdf>
2. Descartes R. Meditations on First Philosophy [Internet]. Broadview Press; 2013. 100 p. Available from: <https://play.google.com/store/books/details?id=Lal2AwAAQBAJ>
3. Muramoto O, Englert WG. Socrates and temporal lobe epilepsy: a pathographic diagnosis 2,400 years later. *Epilepsia* [Internet]. 2006 Mar;47(3):652–4. Available from: <http://dx.doi.org/10.1111/j.1528-1167.2006.00481.x>
4. Cox LA Jr. Causal reasoning about epidemiological associations in conversational AI. *Glob Epidemiol* [Internet]. 2023 Dec;5:100102. Available from: <http://dx.doi.org/10.1016/j.gloepi.2023.100102>
5. Paul RW, Elder L. Critical Thinking: The Nature of Critical and Creative Thought. *Journal of Developmental Education* [Internet]. 2006 [cited 2024 Feb 24]; Available from: <https://www.semanticscholar.org/paper/8bc9e8bfe26e71fdf1cb68c93d7561c478d7c032>
6. Pearl J. Causality: Models, Reasoning and Inference [Internet]. Cambridge University Press; 2009. Available from: <https://play.google.com/store/books/details?id=6bIRkgEACAAJ>
7. Begbie JW. Hippocrates: His Life and Writings. *Br Med J* [Internet]. 1872 Dec 21;2(625):674–7. Available from: <http://dx.doi.org/10.1136/bmj.2.625.674>
8. Kleisiaris CF, Sfakianakis C, Papathanasiou IV. Health care practices in ancient Greece: The Hippocratic ideal. *J Med Ethics Hist Med* [Internet]. 2014 Mar 15;7:6. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/25512827>
9. Salute, Longevita. História da Medicina Integrativa [Internet]. Salute & Longevita. 2018 [cited 2024 Feb 26]. Available from: <https://saluteelongevita.com.br/historia-da-medicina-integrativa/>
10. Cairus HF, Ribeiro WA Jr. Textos hipocráticos: o doente, o médico e a doença [Internet]. SciELO - Editora FIOCRUZ; 2005. 252 p. Available from: <https://play.google.com/store/books/details?id=tlltAwAAQBAJ>
11. Fowler RC. On the Heart of the Hippocratic Corpus: its meaning, context and purpose. *Med Hist* [Internet]. 2023 Jul;67(3):266–83. Available from: <http://dx.doi.org/10.1017/mdh.2023.22>
12. The Extant Works of Aretæus the Cappadocian. *Br Foreign Med Chir Rev* [Internet]. 1857 Apr;19(38):374–81. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/30163973>
13. Reed JA. Aretæus, the Cappadocian: history enlightens the present. *Diabetes* [Internet]. 1954 Sep-Oct;3(5):419–21. Available from: <http://dx.doi.org/10.2337/diab.3.5.419>
14. Howell TH. Aretæus on disease in old age. *J Am Geriatr Soc* [Internet]. 1971 Nov;19(11):909–12. Available from: <http://dx.doi.org/10.1111/j.1532-5415.1971.tb02206.x>

15. Pearce JMS. Early contribution of Alexandria medical school to the anatomy, physiology and pathology of the nervous system. *Rev Neurol* [Internet]. 2019 Mar;175(3):119–25. Available from: <http://dx.doi.org/10.1016/j.neurol.2018.04.011>
16. Laios K, Lytsikas-Sarlis P, Manes K, Kontaxaki MI, Karamanou M, Androutsos G. Drugs for mental illnesses in ancient greek medicine. *Psychiatrike* [Internet]. 2019 Jan-Mar;30(1):58–65. Available from: <http://dx.doi.org/10.22365/jpsych.2019.301.58>
17. Adams F, Aretaeus F. The extant works of aretaeus, the cappadocian [Internet]. Arkose Press; 2015 [cited 2024 Feb 24]. 568 p. Available from: <https://daedalus.umkc.edu/hippocrates/aretaeusEnglish/page.166.a.php?size=240x320>
18. The Extant Works of Aretaeus, The Cappadocian [Internet]. [cited 2024 Feb 24]. Available from: <https://www.chlt.org/hippocrates/aretaeusEnglish/page.1.a.php?size=240x320>
19. Kotsopoulos S. Aretaeus the Cappadocian on mental illness. *Compr Psychiatry* [Internet]. 1986 Mar-Apr;27(2):171–9. Available from: [http://dx.doi.org/10.1016/0010-440x\(86\)90026-x](http://dx.doi.org/10.1016/0010-440x(86)90026-x)
20. Brandt L, Goerig M. [The history of tracheotomy. I]. *Anaesthetist* [Internet]. 1986 May;35(5):279–83. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/3526969>
21. Tekiner H. Figure 2: Engraved portrait of Aretaeus by Ioannes Sambucus (1574) [Internet]. ResearchGate. [cited 2024 Mar 3]. Available from: https://www.researchgate.net/figure/Engraved-portrait-of-Aretaeus-by-Ioannes-Sambucus-1574-Courtesy-of-the-Bibliothèque_fig2_277782979
22. Charles D. Book Review. *Philos Q* [Internet]. 1981 Jul 1 [cited 2024 Feb 24];31(124):269–71. Available from: <https://academic.oup.com/pq/article-lookup/doi/10.2307/2219117>
23. Holland PW. Statistics and Causal Inference. *J Am Stat Assoc* [Internet]. 1986 Dec 1;81(396):945–60. Available from: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354>
24. Hankinson RJ. *Cause and Explanation in Ancient Greek Thought* [Internet]. Clarendon; 1998. 499 p. Available from: <https://play.google.com/store/books/details?id=tqsGvgEACAAJ>
25. Fisher RA. *The design of experiments*. The design of experiments [Internet]. 1960; Available from: <https://www.cabdirect.org/cabdirect/abstract/19601604515>
26. Pearl J. Statistics and causality: separated to reunite-commentary on Bryan Dowd’s “separated at birth.” *Health Serv Res* [Internet]. 2011 Apr;46(2):421–9. Available from: <http://dx.doi.org/10.1111/j.1475-6773.2011.01243.x>
27. Dowd BE. Separated at birth: statisticians, social scientists, and causality in health services research. *Health Serv Res* [Internet]. 2011 Apr;46(2):397–420. Available from: <http://dx.doi.org/10.1111/j.1475-6773.2010.01203.x>
28. Cox DR. Causality: Some statistical aspects. *J R Stat Soc Ser A Stat Soc* [Internet]. 1992;155(2):291. Available from: <https://www.jstor.org/stable/2982962?origin=crossref>
29. Barnard GA. Causation—I. In: *Encyclopedia of Statistical Sciences* [Internet]. Hoboken,

- NJ, USA: John Wiley & Sons, Inc.; 2006. Available from:
<https://onlinelibrary.wiley.com/doi/10.1002/0471667196.ess0209.pub2>
30. Pearson K, Harris JA, Treloar AE, Wilder M. On the Theory of Contingency. *J Am Stat Assoc* [Internet]. 1930;25(171):320–7. Available from:
<http://www.jstor.org/stable/2278196>
 31. Kroese DP, Botev Z, Taimre T, Vaisman R. *Data Science and Machine Learning: Mathematical and Statistical Methods* [Internet]. CRC Press; 2019. 532 p. Available from: <https://play.google.com/store/books/details?id=F7zADwAAQBAJ>
 32. Pearson K. V. On the methemathical theory of errors of judgement, with special reference to the personal equation. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character* [Internet]. 1997 Jan 1;198(300-311):235–99. Available from: <https://doi.org/10.1098/rsta.1902.0005>
 33. Lind J. *A Treatise on the Scurvy: In Three Parts. Containing an Inquiry Into the Nature, Causes, and Cure, of that Disease. Together with a Critical and Chronological View of what Has Been Published on the Subject* [Internet]. S. Crowder; 1772. 559 p. Available from: <https://play.google.com/store/books/details?id=T1OT3tYmh5wC>
 34. Simon HB. *The Harvard Medical School Guide to Men's Health: Lessons from the Harvard Men's Health Studies* [Internet]. Simon and Schuster; 2004. 496 p. Available from: https://play.google.com/store/books/details?id=ZS_OxSy11EsC
 35. Dan. James Lind [Internet]. *Missing the Forest for the Tree: A Worldview Grounded in Science*. Missing The Forest; 2019 [cited 2024 Mar 3]. Available from:
<https://www.missingtheforest.com/james-lind/>
 36. Morabia A. Pierre-Charles-Alexandre Louis and the evaluation of bloodletting. *J R Soc Med* [Internet]. 2006 Mar;99(3):158–60. Available from:
<http://dx.doi.org/10.1177/014107680609900322>
 37. 28th Bradford hill memorial lecture [Internet]. LSHTM. [cited 2024 Mar 4]. Available from:
<https://www.lshtm.ac.uk/newsevents/events/28th-bradford-hill-memorial-lecture>
 38. Ronald A. Fisher [Internet]. 42. 2014 [cited 2024 Mar 3]. Available from:
<https://www.42evolution.org/ronald-a-fisher/>
 39. Armitage P. Fisher, Bradford Hill, and randomization. *Int J Epidemiol* [Internet]. 2003 Dec;32(6):925–8; discussion 945–8. Available from: <http://dx.doi.org/10.1093/ije/dyg286>
 40. Fisher SRA. Smoking: the cancer controversy: some attempts to assess the evidence [Internet]. [cited 2024 Feb 24]. Available from:
https://lru.praxis.dk/Lru/microsites/hvadermatematik/hem2download/kap10_projekt_10_2_ekstra_FisherOnSmokingAndCancer_samlet.pdf
 41. Burch PR. Cigarette smoking and lung cancer: a continuing controversy. *Med Hypotheses* [Internet]. 1982 Sep;9(3):293–306. Available from:
[http://dx.doi.org/10.1016/0306-9877\(82\)90158-x](http://dx.doi.org/10.1016/0306-9877(82)90158-x)
 42. Christopher B. Why the father of modern statistics didn't believe smoking caused cancer [Internet]. *Priceonomics*. 2016 [cited 2024 Feb 24]. Available from:
<https://priceonomics.com/why-the-father-of-modern-statistics-didnt-believe/>

43. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *Br Med J* [Internet]. 1950 Sep 30;2(4682):739–48. Available from: <http://dx.doi.org/10.1136/bmj.2.4682.739>
44. Di Cicco ME, Ragazzo V, Jacinto T. Mortality in relation to smoking: the British Doctors Study. *Breathe (Sheff)* [Internet]. 2016 Sep;12(3):275–6. Available from: <http://dx.doi.org/10.1183/20734735.013416>
45. Burch PR. Smoking, lung cancer and hypothesis testing. *Med Hypotheses* [Internet]. 1981 Dec;7(12):1461–70. Available from: [http://dx.doi.org/10.1016/0306-9877\(81\)90145-6](http://dx.doi.org/10.1016/0306-9877(81)90145-6)
46. Akobeng AK. Understanding randomised controlled trials. *Arch Dis Child* [Internet]. 2005 Aug;90(8):840–4. Available from: <http://dx.doi.org/10.1136/adc.2004.058222>
47. Pearl J, Mackenzie D. The book of why: the new science of cause and effect [Internet]. Basic books; 2018. Available from: http://repo.darmajaya.ac.id/4847/1/The%20book%20of%20why_%20the%20new%20science%20of%20cause%20and%20effect%20%28%20PDFDrive%20%29.pdf
48. French GL, Cheng AF, Wong SL, Donnan S. Repeated prevalence surveys for monitoring effectiveness of hospital infection control. *Lancet* [Internet]. 1989 Oct 28;2(8670):1021–3. Available from: [http://dx.doi.org/10.1016/s0140-6736\(89\)91026-x](http://dx.doi.org/10.1016/s0140-6736(89)91026-x)
49. Ministério da Saúde [Internet]. [cited 2024 Feb 24]. Available from: https://bvsmis.saude.gov.br/bvs/saudelegis/gm/1998/prt2616_12_05_1998.html
50. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care–associated infection and criteria for specific types of infections in the acute care setting. *Am J Infect Control* [Internet]. 2008 Jun 1;36(5):309–32. Available from: <https://doi.org/10.1016/j.ajic.2008.03.002>
51. Data Portal [Internet]. 2023 [cited 2024 Feb 24]. Available from: <https://www.cdc.gov/hai/data/portal/index.html>
52. Health Organization W. Global health risks: mortality and burden of disease attributable to selected major risks [Internet]. [cited 2022 Nov 15]. Available from: https://apps.who.int/iris/bitstream/handle/10665/44203/9789241563871_eng.pdf
53. Research Committee of the Society of Healthcare Epidemiology of America. Enhancing patient safety by reducing healthcare-associated infections: the role of discovery and dissemination. *Infect Control Hosp Epidemiol* [Internet]. 2010 Feb;31(2):118–23. Available from: <http://dx.doi.org/10.1086/650198>
54. Agência Nacional de Vigilância Sanitária [Internet]. [cited 2024 Feb 24]. Available from: <https://www.gov.br/pt-br/orgaos/agencia-nacional-de-vigilancia-sanitaria>
55. Anvisa: qual o papel desse órgão na saúde do Brasil? [Internet]. Politize!; 2021 [cited 2024 Feb 24]. Available from: <https://www.politize.com.br/anvisa-na-saude-do-brasil/>
56. Segurança do paciente, gestão de riscos e controle de infecções hospitalares [Internet]. *Issuu*. 2016 [cited 2024 Feb 24]. Available from: https://issuu.com/moriaeditoraltda/docs/seguran__a_do_paciente__gest__o_de__
57. Hebden JN, Wright MO, Fuss EP, Standiford HC. Leveraging surveillance technology to

- benefit the practice and profession of infection control. *Am J Infect Control* [Internet]. 2008 Apr 1;36(3, Supplement):S7–11. Available from: <https://www.sciencedirect.com/science/article/pii/S0196655307007808>
58. Ministério da Saúde [Internet]. [cited 2024 Feb 24]. Available from: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2017/prt2436_22_09_2017.html
 59. Puro V, Coppola N, Frasca A, Gentile I, Luzzaro F, Peghetti A, et al. Pillars for prevention and control of healthcare-associated infections: an Italian expert opinion statement. *Antimicrob Resist Infect Control* [Internet]. 2022 Jun 20;11(1):87. Available from: <http://dx.doi.org/10.1186/s13756-022-01125-8>
 60. Strengthening National Efforts to Reduce Healthcare-Associated Infections [Internet]. [cited 2024 Feb 24]. Available from: <https://www.ahrq.gov/hai/patient-safety-resources/advances-in-hai/hai-article2.html>
 61. Furuno JP, Schweizer ML, McGregor JC, Perencevich EN. Economics of infection control surveillance technology: cost-effective or just cost? *Am J Infect Control* [Internet]. 2008 Apr;36(3 Suppl):S12–7. Available from: <http://dx.doi.org/10.1016/j.ajic.2007.06.002>
 62. Glenister HM, Taylor LJ, Bartlett CL, Cooke EM, Sedgwick JA, Mackintosh CA. An evaluation of surveillance methods for detecting infections in hospital inpatients. *J Hosp Infect* [Internet]. 1993 Mar;23(3):229–42. Available from: [http://dx.doi.org/10.1016/0195-6701\(93\)90028-x](http://dx.doi.org/10.1016/0195-6701(93)90028-x)
 63. Araújo BT, Pereira DCR. Policies for control Health Care-Related Infections (IRAS) in Brazil, 2017. *IEEE Internet Comput*. 2017;28(3/4):333–42.
 64. Bearman G, Doll M, Cooper K, Stevens MP. Hospital Infection Prevention: How Much Can We Prevent and How Hard Should We Try? *Curr Infect Dis Rep* [Internet]. 2019 Feb 2;21(1):2. Available from: <http://dx.doi.org/10.1007/s11908-019-0660-2>
 65. Iancu D, Moldovan I, Țilea B, Voidăzan S. Evaluating Healthcare-Associated Infections in Public Hospitals: A Cross-Sectional Study. *Antibiotics (Basel)* [Internet]. 2023 Dec 2;12(12). Available from: <http://dx.doi.org/10.3390/antibiotics12121693>
 66. Stacy KM. Challenges in Hospital-Associated Infection Management: A Unit Perspective. *AACN Adv Crit Care* [Internet]. 2015 Jul-Sep;26(3):252–61. Available from: <http://dx.doi.org/10.1097/NCI.0000000000000097>
 67. Williamson L. The ongoing challenge of managing and preventing hospital-acquired infections. *Br J Health Care Manag* [Internet]. 2022 Jan 2;28(1):26–9. Available from: <https://www.sciencegate.app/document/10.12968/bjhc.2021.0137>
 68. Nicolau DP. Current challenges in the management of the infected patient. *Curr Opin Infect Dis* [Internet]. 2011 Feb;24 Suppl 1:S1–10. Available from: <http://dx.doi.org/10.1097/01.qco.0000393483.10270.ff>
 69. Haque M, McKimm J, Sartelli M, Dhingra S, Labricciosa FM, Islam S, et al. Strategies to Prevent Healthcare-Associated Infections: A Narrative Overview. *Risk Manag Healthc Policy* [Internet]. 2020 Sep 28;13:1765–80. Available from: <http://dx.doi.org/10.2147/RMHP.S269315>
 70. Bates DW, Levine D, Syrowatka A, Kuznetsova M, Craig KJT, Rui A, et al. The potential

- of artificial intelligence to improve patient safety: a scoping review. *NPJ Digit Med* [Internet]. 2021 Mar 19;4(1):54. Available from: <http://dx.doi.org/10.1038/s41746-021-00423-6>
71. Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A. Artificial intelligence-based tools to control healthcare associated infections: A systematic review of the literature. *J Infect Public Health* [Internet]. 2020 Aug;13(8):1061–77. Available from: <http://dx.doi.org/10.1016/j.jiph.2020.06.006>
 72. Dos Santos RP, Silva D, Menezes A, Lukasewicz S, Dalmora CH, Carvalho O, et al. Automated healthcare-associated infection surveillance using an artificial intelligence algorithm. *Infect Prev Pract* [Internet]. 2021 Sep;3(3):100167. Available from: <http://dx.doi.org/10.1016/j.infpip.2021.100167>
 73. Lukasewicz Ferreira Stephani Amanda, Franco Meneses Arateus Crysham, Vaz Tiago Andres, da Fontoura Carvalho Otávio Luiz, Dalmora Camila Hubner, Vanni Daiane Pressoto, et al. Using Machine Learning to Reduce Burden on Infection Control Staff. *NEJM Catalyst* [Internet]. 3(8):CAT.22.0071. Available from: <https://doi.org/10.1056/CAT.22.0071>
 74. Ambika P. Chapter Thirteen - Machine learning and deep learning algorithms on the Industrial Internet of Things (IIoT). In: Raj P, Evangeline P, editors. *Advances in Computers* [Internet]. Elsevier; 2020. p. 321–38. Available from: <https://www.sciencedirect.com/science/article/pii/S0065245819300609>
 75. Painuli D, Mishra D, Bhardwaj S, Aggarwal M. 20 - Forecast and prediction of COVID-19 using machine learning. In: Kose U, Gupta D, de Albuquerque VHC, Khanna A, editors. *Data Science for COVID-19* [Internet]. Academic Press; 2021. p. 381–97. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128245361000277>
 76. An Q, Rahman S, Zhou J, Kang JJ. A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors* [Internet]. 2023 Apr 22;23(9). Available from: <http://dx.doi.org/10.3390/s23094178>
 77. Ambinder EP. Electronic health records. *J Oncol Pract* [Internet]. 2005 Jul;1(2):57–63. Available from: <http://dx.doi.org/10.1200/JOP.2005.1.2.57>
 78. Alzu'bi AA, Watzlaf VJM, Sheridan P. Electronic Health Record (EHR) Abstraction. *Perspect Health Inf Manag* [Internet]. 2021 Mar 15;18(Spring):1g. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/34035788>
 79. Singh S, Tripathi BK. Pneumonia classification using quaternion deep learning. *Multimed Tools Appl* [Internet]. 2022;81(2):1743–64. Available from: <http://dx.doi.org/10.1007/s11042-021-11409-7>
 80. Slagter JS, Outmani L, Tran KTCK, Ijzermans JNM, Minnee RC. Robot-assisted kidney transplantation as a minimally invasive approach for kidney transplant recipients: A systematic review and meta-analyses. *Int J Surg* [Internet]. 2022 Mar;99:106264. Available from: <http://dx.doi.org/10.1016/j.ijsu.2022.106264>
 81. Galatzan BJ, Carrington JM, Gephart S. Testing the Use of Natural Language Processing Software and Content Analysis to Analyze Nursing Hand-off Text Data. *Comput Inform Nurs* [Internet]. 2021 May 10;39(8):411–7. Available from: <http://dx.doi.org/10.1097/CIN.0000000000000732>

82. Schneider ETR, de Souza JVA, Knafou J, Oliveira LES e., Copara J, Gumiel YB, et al. BioBERTpt - A Portuguese Neural Language Model for Clinical Named Entity Recognition. In: Proceedings of the 3rd Clinical Natural Language Processing Workshop [Internet]. Online: Association for Computational Linguistics; 2020. p. 65–72. Available from: <https://www.aclweb.org/anthology/2020.clinicalnlp-1.7>
83. Tang R, Han X, Jiang X, Hu X. Does Synthetic Data Generation of LLMs Help Clinical Text Mining? [Internet]. arXiv [cs.CL]. 2023. Available from: <http://arxiv.org/abs/2303.04360>
84. Li Z, Zhu H, Lu Z, Yin M. Synthetic Data Generation with Large Language Models for Text Classification: Potential and Limitations [Internet]. arXiv [cs.CL]. 2023. Available from: <http://arxiv.org/abs/2310.07849>
85. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. *Commun Med* [Internet]. 2023 Oct 10;3(1):141. Available from: <http://dx.doi.org/10.1038/s43856-023-00370-1>
86. Xie Q, Schenck EJ, Yang HS, Chen Y, Peng Y, Wang F. Faithful AI in Medicine: A Systematic Review with Large Language Models and Beyond. *medRxiv* [Internet]. 2023 Jul 1; Available from: <http://dx.doi.org/10.1101/2023.04.18.23288752>
87. Naucier P, van der Werff SD, Valik J, Ward L, Ternhag A, Tanushi H, et al. HAI-Proactive: Development of an Automated Surveillance System for Healthcare-Associated Infections in Sweden. *Infect Control Hosp Epidemiol* [Internet]. 2020 Oct [cited 2024 Feb 24];41(S1):s39–s39. Available from: <https://www.cambridge.org/core/journals/infection-control-and-hospital-epidemiology/article/haiproactive-development-of-an-automated-surveillance-system-for-healthcare-associated-infections-in-sweden/5656247BF5849125268DE21E38939318>
88. Gubbels S, Nielsen J, Voldstedlund M, Kristensen B, Schönheyder HC, Ellermann-Eriksen S, et al. National Automated Surveillance of Hospital-Acquired Bacteremia in Denmark Using a Computer Algorithm. *Infect Control Hosp Epidemiol* [Internet]. 2017 May;38(5):559–66. Available from: <http://dx.doi.org/10.1017/ice.2017.1>
89. Affonso D. Entenda como funciona o robô usado pelo Tacchini, de Bento, para identificação de infecções [Internet]. GZH. 2021 [cited 2024 Feb 26]. Available from: <https://gauchazh.clicrbs.com.br/pioneiro/geral/noticia/2021/11/entenda-como-funciona-o-robo-usado-pelo-tacchini-de-bento-para-identificacao-de-infecoes-ckvjtdyd70082019m28bmezdh.html>
90. Verberk JDM, van der Kooi TII, Hetem DJ, Oostdam NEWM, Noordergraaf M, de Greeff SC, et al. Semiautomated surveillance of deep surgical site infections after colorectal surgeries: A multicenter external validation of two surveillance algorithms. *Infect Control Hosp Epidemiol* [Internet]. 2023 Apr;44(4):616–23. Available from: <http://dx.doi.org/10.1017/ice.2022.147>
91. Healthcare LLM [Internet]. John Snow Labs. 2023 [cited 2024 Feb 26]. Available from: <https://www.johnsnowlabs.com/healthcare-llm/>
92. Healthcare NLP [Internet]. John Snow Labs. 2023 [cited 2024 Feb 26]. Available from: <https://www.johnsnowlabs.com/healthcare-nlp/>
93. Verberk JDM, Aghdassi SJS, Abbas M, Naucier P, Gubbels S, Maldonado N, et al.

- Automated surveillance systems for healthcare-associated infections: results from a European survey and experiences from real-life utilization. *J Hosp Infect* [Internet]. 2022 Apr;122:35–43. Available from: <http://dx.doi.org/10.1016/j.jhin.2021.12.021>
94. Chiu HYR, Hwang CK, Chen SY, Shih FY, Han HC, King CC, et al. Machine learning for emerging infectious disease field responses. *Sci Rep* [Internet]. 2022 Jan 10;12(1):328. Available from: <http://dx.doi.org/10.1038/s41598-021-03687-w>
 95. Cho SY, Kim Z, Chung DR, Cho BH, Chung MJ, Kim JH, et al. Development of machine learning models for the surveillance of colon surgical site infections. *J Hosp Infect* [Internet]. 2023 Apr 22; Available from: <http://dx.doi.org/10.1016/j.jhin.2023.03.025>
 96. Theodosiou AA, Read RC. Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician. *J Infect* [Internet]. 2023 Oct;87(4):287–94. Available from: <http://dx.doi.org/10.1016/j.jinf.2023.07.006>
 97. Sips ME, Bonten MJM, van Mourik MSM. Automated surveillance of healthcare-associated infections: state of the art. *Curr Opin Infect Dis* [Internet]. 2017 Aug;30(4):425–31. Available from: <http://dx.doi.org/10.1097/QCO.0000000000000376>
 98. Wright MO. Automated surveillance and infection control: Toward a better tomorrow. *Am J Infect Control* [Internet]. 2008 Apr 1;36(3, Supplement):S1–6. Available from: <https://www.sciencedirect.com/science/article/pii/S019665530700689X>
 99. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swami A. The Limitations of Deep Learning in Adversarial Settings. In: 2016 IEEE European Symposium on Security and Privacy (EuroS&P) [Internet]. 2016. p. 372–87. Available from: <http://dx.doi.org/10.1109/EuroSP.2016.36>
 100. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG* [Internet]. 2018 Dec;125(13):1716. Available from: <http://dx.doi.org/10.1111/1471-0528.15199>
 101. Westreich D. *Epidemiology by Design: A Causal Approach to the Health Sciences* [Internet]. Oxford University Press; 2019. 288 p. Available from: <https://play.google.com/store/books/details?id=IEK1DwAAQBAJ>
 102. Brown KE, Flores MJ, Slobogean G, Shearer D, Gitajn IL, Morshed S. Simple design and analysis strategies for solving problems in observational orthopaedic clinical research. *OTA Int* [Internet]. 2023 May;6(2 Suppl):e239. Available from: <http://dx.doi.org/10.1097/OI9.0000000000000239>
 103. Hume D. *An Enquiry Concerning the Human Understanding: And An Enquiry Concerning the Principles of Morals* [Internet]. Clarendon Press; 1894. 349 p. Available from: <https://play.google.com/store/books/details?id=DTU7AAAAYAAJ>
 104. Greenland S, Pearl J, Robins JM. Confounding and Collapsibility in Causal Inference. *SSO Schweiz Monatsschr Zahnheilkd* [Internet]. 1999 Feb [cited 2023 Jun 21];14(1):29–46. Available from: <https://projecteuclid.org/journals/statistical-science/volume-14/issue-1/Confounding-and-Collapsibility-in-Causal-Inference/10.1214/ss/1009211805.short>
 105. Pearl J. An introduction to causal inference. *Int J Biostat* [Internet]. 2010 Feb 26;6(2):Article 7. Available from: <http://dx.doi.org/10.2202/1557-4679.1203>

106. Hernán MA, Robins JM. Causal Inference: What If [Internet]. Chapman & Hall/CRC; 331 p. Available from: <https://play.google.com/store/books/details?id=mOh7zwEACAAJ>
107. Msaouel P, Lee J, Karam JA, Thall PF. A Causal Framework for Making Individualized Treatment Decisions in Oncology. *Cancers* [Internet]. 2022 Aug 14;14(16). Available from: <http://dx.doi.org/10.3390/cancers14163923>
108. Pearl J, Bareinboim E. External Validity: From Do-Calculus to Transportability Across Populations. In: *Probabilistic and Causal Inference: The Works of Judea Pearl* [Internet]. 1st ed. New York, NY, USA: Association for Computing Machinery; 2022 [cited 2024 Feb 23]. p. 451–82. Available from: <https://doi.org/10.1145/3501714.3501741>
109. Shapiro DD, Msaouel P. Causal Diagram Techniques for Urologic Oncology Research. *Clin Genitourin Cancer* [Internet]. 2021 Jun;19(3):271.e1–271.e7. Available from: <http://dx.doi.org/10.1016/j.clgc.2020.08.003>
110. Hufstedler H, Rahman S, Danzer AM, Goymann H, de Jong VMT, Campbell H, et al. Systematic Review Reveals Lack of Causal Methodology Applied to Pooled Longitudinal Observational Infectious Disease Studies. *J Clin Epidemiol* [Internet]. 2022 May;145:29–38. Available from: <http://dx.doi.org/10.1016/j.jclinepi.2022.01.008>
111. Sharma N, Srivastav A, Samuel AJ. Randomized clinical trial: gold standard of experimental designs - importance, advantages, disadvantages and prejudices. *Rev Pesqui Em Fisioter* [Internet]. 2020 Aug 26; Available from: <https://www5.bahiana.edu.br/index.php/fisioterapia/article/view/3039>
112. Glymour MM. Commentary: Modelling change in a causal framework. *Int J Epidemiol* [Internet]. 2022 Oct 13;51(5):1615–21. Available from: <http://dx.doi.org/10.1093/ije/dyab151>
113. Tennant PWG, Arnold KF, Ellison GTH, Gilthorpe MS. Analyses of “change scores” do not estimate causal effects in observational data. *Int J Epidemiol* [Internet]. 2022 Oct 13;51(5):1604–15. Available from: <http://dx.doi.org/10.1093/ije/dyab050>
114. Halloran ME, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology* [Internet]. 1995 Mar;6(2):142–51. Available from: <http://dx.doi.org/10.1097/00001648-199503000-00010>
115. Paynter S. Incorporating Transmission Into Causal Models of Infectious Diseases for Improved Understanding of the Effect and Impact of Risk Factors. *Am J Epidemiol* [Internet]. 2016 Mar 15;183(6):574–82. Available from: <http://dx.doi.org/10.1093/aje/kwv234>
116. Cartwright N. Are RCTs the Gold Standard? *Biosocieties* [Internet]. 2007 Mar [cited 2024 Feb 24];2(1):11–20. Available from: <https://www.cambridge.org/core/journals/biosocieties/article/abs/are-rcts-the-gold-standard/0A53C381EDDC6FE52EF8D5E79B2A179B>
117. Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. *J Eval Clin Pract* [Internet]. 2010 Apr;16(2):260–6. Available from: <http://dx.doi.org/10.1111/j.1365-2753.2010.01382.x>
118. Port FK. Role of observational studies versus clinical trials in ESRD research. *Kidney Int* [Internet]. 2000 Jan 1 [cited 2024 Feb 24];57:S3–6. Available from:

<http://www.kidney-international.org/article/S0085253815470346/abstract>

119. Hufstедler H, Matthey EC, Rahman S, de Jong VMT, Campbell H, Gustafson P, et al. Current trends in the application of causal inference methods to pooled longitudinal observational infectious disease studies-A protocol for a methodological systematic review. *PLoS One* [Internet]. 2021 Apr 29;16(4):e0250778. Available from: <http://dx.doi.org/10.1371/journal.pone.0250778>
120. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* [Internet]. 2005;95 Suppl 1:S144–50. Available from: <http://dx.doi.org/10.2105/AJPH.2004.059204>
121. Pearl J, Glymour M, Jewell NP. *Causal Inference in Statistics: A Primer* [Internet]. John Wiley & Sons; 2016. 160 p. Available from: <https://play.google.com/store/books/details?id=I0V2CwAAQBAJ>
122. Pearl J, Verma TS. A theory of inferred causation. In: Prawitz D, Skyrms B, Westerståhl D, editors. *Studies in Logic and the Foundations of Mathematics* [Internet]. Elsevier; 1995. p. 789–811. Available from: <https://www.sciencedirect.com/science/article/pii/S0049237X06800741>
123. Elwert F. Graphical Causal Models. In: *Handbook of Causal Analysis for Social Research* [Internet]. unknown; 2013 [cited 2024 Feb 24]. p. 245–73. Available from: https://www.researchgate.net/publication/278717528_Graphical_Causal_Models
124. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* [Internet]. 1999 Jan;10(1):37–48. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/9888278>
125. Beretta S, Castelli M, Gonçalves I, Henriques R, Ramazzotti D. Learning the Structure of Bayesian Networks: A Quantitative Assessment of the Effect of Different Algorithmic Schemes. *Complexity* [Internet]. 2018 Sep 12 [cited 2024 Feb 24];2018. Available from: <https://www.hindawi.com/journals/complexity/2018/1591878/>
126. Verma T, Pearl J. Causal Networks: Semantics and Expressiveness* *This work was partially supported by the National Science Foundation Grant #IRI-8610155. “Graphoids: A Computer Representation for Dependencies and Relevance in Automated Reasoning (Computer Information Science).” In: Shachter RD, Levitt TS, Kanal LN, Lemmer JF, editors. *Machine Intelligence and Pattern Recognition* [Internet]. North-Holland; 1990. p. 69–76. Available from: <https://www.sciencedirect.com/science/article/pii/B9780444886507500111>
127. Greenland S, Pearl J. Causal Diagrams. 2007 [cited 2024 Feb 24]; Available from: <https://escholarship.org/uc/item/7tn3p6jx>
128. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* [Internet]. 1999 Jan;10(1):37–48. Available from: <https://escholarship.org/content/qt7tn3p6jx/qt7tn3p6jx.pdf>
129. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* [Internet]. 2002 Jan 15;155(2):176–84. Available from: <http://dx.doi.org/10.1093/aje/155.2.176>

130. Pearl J. Causal Inference. In: Guyon I, Janzing D, Schölkopf B, editors. Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008 [Internet]. Whistler, Canada: PMLR; 2010. p. 39–58. (Proceedings of Machine Learning Research; vol. 6). Available from: <https://proceedings.mlr.press/v6/pearl10a.html>
131. Rohrer JM. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science* [Internet]. 2018 Mar 1;1(1):27–42. Available from: <https://doi.org/10.1177/2515245917745629>
132. Glymour MM, Greenland S. Causal diagrams. *Modern epidemiology* [Internet]. 2008; Available from: <http://publicifsv.sund.ku.dk/~jhp/KandidatFSV/aar19/Forelaesninger/DAG/GlymourChap12.pdf>
133. Garnier M, Constantin JM, Heming N, Camous L, Ferré A, Razazi K, et al. Epidemiology, risk factors and prognosis of ventilator-associated pneumonia during severe COVID-19: Multicenter observational study across 149 European Intensive Care Units. *Anaesth Crit Care Pain Med* [Internet]. 2023 Feb;42(1):101184. Available from: <http://dx.doi.org/10.1016/j.accpm.2022.101184>
134. Pearl J. Causal diagrams for empirical research. *Biometrika* [Internet]. 1995 Dec 1 [cited 2024 Feb 24];82(4):669–88. Available from: <https://academic.oup.com/biomet/article-abstract/82/4/669/251647>
135. Pearl J. Graphical models, causality, and intervention [Internet]. [cited 2024 Feb 25]. Available from: <https://escholarship.org/content/qt8d93w51g/qt8d93w51g.pdf>
136. Causal Inference The Mixtape - 3 Directed Acyclic Graphs [Internet]. [cited 2024 Feb 24]. Available from: https://mixtape.scunning.com/03-directed_acyclical_graphs
137. Carmona-Bayonas A, Jiménez-Fonseca P, Gallego J, Msaouel P. Causal Considerations Can Inform the Interpretation of Surprising Associations in Medical Registries. *Cancer Invest* [Internet]. 2022 Jan;40(1):1–13. Available from: <http://dx.doi.org/10.1080/07357907.2021.1999971>
138. Msaouel P. The Big Data Paradox in Clinical Practice. *Cancer Invest* [Internet]. 2022 Aug;40(7):567–76. Available from: <http://dx.doi.org/10.1080/07357907.2022.2084621>
139. Causal Inference from Observational Data [Internet]. Miguel Hernan's Faculty Website. 2018 [cited 2024 Feb 24]. Available from: <https://www.hsph.harvard.edu/miguel-hernan/research/causal-inference-from-observational-data/>
140. Olier I, Zhan Y, Liang X, Volovici V. Causal inference and observational data. *BMC Med Res Methodol* [Internet]. 2023 Oct 11;23(1):227. Available from: <http://dx.doi.org/10.1186/s12874-023-02058-5>
141. Spirtes P, Glymour CN, Scheines R. *Causation, Prediction, and Search* [Internet]. MIT Press; 2000. 543 p. Available from: <https://play.google.com/store/books/details?id=vV-U09kCdRwC>
142. Pearl J. Fusion, Propagation, and Structuring in Belief Networks. In: *Probabilistic and Causal Inference: The Works of Judea Pearl* [Internet]. 1st ed. New York, NY, USA:

- Association for Computing Machinery; 2022 [cited 2024 Feb 24]. p. 139–88. Available from: <https://doi.org/10.1145/3501714.3501728>
143. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference [Internet]. Morgan Kaufmann; 1988. 552 p. Available from: <https://play.google.com/store/books/details?id=AvNID7LyMusC>
 144. Pearl J. [Bayesian Analysis in Expert Systems]: Comment: Graphical Models, Causality and Intervention. *Stat Sci* [Internet]. 1993;8(3):266–9. Available from: <http://www.jstor.org/stable/2245965>
 145. Sadeghi K, Soo T. Conditions and assumptions for constraint-based causal structure learning. *J Mach Learn Res* [Internet]. 2022 Jan 1;23(1):4699–732. Available from: <https://dl.acm.org/doi/abs/10.5555/3586589.3586698>
 146. Peters J, Janzing D, Schölkopf B. Elements of causal inference: foundations and learning algorithms [Internet]. The MIT Press; 2017. Available from: <https://library.oapen.org/handle/20.500.12657/26040>
 147. Neuberger LG. CAUSALITY: MODELS, REASONING, AND INFERENCE, by Judea Pearl, Cambridge University Press, 2000. *Econometric Theory* [Internet]. 2003 Aug [cited 2024 Feb 25];19(4):675–85. Available from: <https://www.cambridge.org/core/journals/econometric-theory/article/causality-models-reasoning-and-inference-by-judea-pearl-cambridge-university-press-2000/DA2D9ABB0AD3DAC95AE7B3081FCDF139>
 148. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci U S A* [Internet]. 2016 Jul 5;113(27):7345–52. Available from: <http://dx.doi.org/10.1073/pnas.1510507113>
 149. Heiss A. Do-calculus adventures! Exploring the three rules of do-calculus in plain language and deriving the backdoor adjustment formula by hand [Internet]. 2021 [cited 2024 Feb 25]. Available from: <https://www.andrewheiss.com/blog/2021/09/07/do-calculus-backdoors/>
 150. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* [Internet]. 2004 Sep;15(5):615–25. Available from: <http://dx.doi.org/10.1097/01.ede.0000135174.63482.43>
 151. Bareinboim E, Pearl J. Transportability of Causal Effects: Completeness Results. *AAAI* [Internet]. 2012 [cited 2024 Feb 25];26(1):698–704. Available from: <https://ojs.aaai.org/index.php/AAAI/article/view/8232>
 152. Bareinboim E, Tian J, Pearl J. Recovering from Selection Bias in Causal and Statistical Inference. In: *Probabilistic and Causal Inference: The Works of Judea Pearl* [Internet]. 1st ed. New York, NY, USA: Association for Computing Machinery; 2022 [cited 2024 Feb 24]. p. 433–50. Available from: <https://doi.org/10.1145/3501714.3501740>
 153. Bareinboim E, Tian J. Recovering Causal Effects from Selection Bias. *AAAI* [Internet]. 2015 Mar 4 [cited 2024 Feb 25];29(1). Available from: <https://ojs.aaai.org/index.php/AAAI/article/view/9679>
 154. Bareinboim E, Tian J, Pearl J. Recovering from Selection Bias in Causal and Statistical Inference. *AAAI* [Internet]. 2014 Jun 21 [cited 2024 Feb 25];28(1). Available

from: <https://ojs.aaai.org/index.php/AAAI/article/view/9074>

155. Correa JD, Tian J, Bareinboim E. Identification of Causal Effects in the Presence of Selection Bias. *AAAI* [Internet]. 2019 Jul 17 [cited 2024 Feb 25];33(01):2744–51. Available from: <https://ojs.aaai.org/index.php/AAAI/article/view/4125>
156. Mohan K, Pearl J. Graphical Models for Processing Missing Data [Internet]. arXiv [stat.ME]. 2018. Available from: <http://arxiv.org/abs/1801.03583>
157. Pearl J. Graphs, Causality, and Structural Equation Models. *Sociol Methods Res* [Internet]. 1998 Nov 1;27(2):226–84. Available from: <https://doi.org/10.1177/0049124198027002004>
158. Höfler M. Causal inference based on counterfactuals. *BMC Med Res Methodol* [Internet]. 2005 Sep 13;5:28. Available from: <http://dx.doi.org/10.1186/1471-2288-5-28>
159. Pearl J. Causal and Counterfactual Inference. In 2018. Available from: <https://api.semanticscholar.org/CorpusID:53061670>
160. Schulz M. Counterfactuals and Probability [Internet]. Oxford University Press; 2017. 256 p. Available from: https://play.google.com/store/books/details?id=nqr_DQAAQBAJ
161. Mulaik SA. Linear Causal Modeling with Structural Equations [Internet]. CRC Press; 2009. 468 p. Available from: <https://play.google.com/store/books/details?id=HgsHnAfVfXEC>
162. Chen B. GRAPHICAL TOOLS FOR LINEAR PATH MODELS. In 2018. Available from: <https://api.semanticscholar.org/CorpusID:4696674>
163. Bollen KA, Pearl J. Eight Myths About Causality and Structural Equation Models. In: Morgan SL, editor. *Handbook of Causal Analysis for Social Research* [Internet]. Dordrecht: Springer Netherlands; 2013. p. 301–28. Available from: https://doi.org/10.1007/978-94-007-6094-3_15
164. Chen B, Pearl J. Graphical Tools for Linear Structural Equation Modeling. In 2014. Available from: <https://api.semanticscholar.org/CorpusID:5452424>
165. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction* [Internet]. Oxford University Press; 2015. 706 p. Available from: <https://play.google.com/store/books/details?id=gBYoBgAAQBAJ>
166. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* [Internet]. 1992 Mar;3(2):143–55. Available from: <http://dx.doi.org/10.1097/00001648-199203000-00013>
167. Bareinboim E, Pearl J. Causal inference by surrogate experiments: Z-identifiability. *Uncertain Artif Intell* [Internet]. 2012 Aug 14;113–20. Available from: <http://arxiv.org/abs/1210.4842>
168. Pearl J. The seven tools of causal inference, with reflections on machine learning. *Commun ACM* [Internet]. 2019 Feb 21;62(3):54–60. Available from: <https://doi.org/10.1145/3241036>
169. HAI checklists [Internet]. 2024 [cited 2024 Feb 25]. Available from: <https://www.cdc.gov/nhsn/hai-checklists/>

170. Romano R, Gambale E. Statistics and medicine: the indispensable know-how of the researcher. *Transl Med UniSa* [Internet]. 2013 Jan;5:28–31. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23905080>
171. Lukasewicz Ferreira SA, Franco Meneses AC, Vaz TA, da Fontoura Carvalho OL, Hubner Dalmora C, Pressotto Vanni D, et al. Hospital-acquired infections surveillance: The machine-learning algorithm mirrors National Healthcare Safety Network definitions. *Infect Control Hosp Epidemiol* [Internet]. 2024 Jan 11;1–5. Available from: <http://dx.doi.org/10.1017/ice.2023.224>
172. Montiroso R, Provenzi L, Tronick E. Risk and Protective Factors Associated with Early Adversity and Development: Evidence from Human and Animal Research [Internet]. *Frontiers Media SA*; 2020. 218 p. Available from: <https://play.google.com/store/books/details?id=3MjPDwAAQBAJ>
173. Sferra SR, Donnelly S, Kabagambe S, Fallon EM. Modifiable Preoperative Risk Factors to Mitigate Postoperative Site Infection Following Pediatric Gastrostomy. *J Pediatr Surg* [Internet]. 2024 Feb 13; Available from: <http://dx.doi.org/10.1016/j.jpedsurg.2024.02.007>
174. Murni IK, Duke T, Kinney S, Daley AJ, Wirawan MT, Soenarto Y. Risk factors for healthcare-associated infection among children in a low-and middle-income country. *BMC Infect Dis* [Internet]. 2022 Apr 26;22(1):406. Available from: <http://dx.doi.org/10.1186/s12879-022-07387-2>
175. Bracken MB. Risk, Chance, and Causation: Investigating the Origins and Treatment of Disease [Internet]. *Yale University Press*; 2013. 330 p. Available from: <https://play.google.com/store/books/details?id=zB4ZHQY5tHYC>
176. Piccininni M, Konigorski S, Rohmann JL, Kurth T. Directed acyclic graphs and causal thinking in clinical risk prediction modeling. *BMC Med Res Methodol* [Internet]. 2020 Jul 2;20(1):179. Available from: <http://dx.doi.org/10.1186/s12874-020-01058-z>
177. Dekkers OM, Vandenbroucke JP. Causaliteit: over risicofactoren en interventies [Internet]. *NTvG*. 2013 [cited 2024 Feb 25]. Available from: https://www.ntvg.nl/artikelen/causaliteit-over-risicofactoren-en-interventies?check_logged_in=1
178. Bellizio F, Cremer JL, Sun M, Strbac G. A causality based feature selection approach for data-driven dynamic security assessment. *Electric Power Systems Research* [Internet]. 2021 Dec 1;201:107537. Available from: <https://www.sciencedirect.com/science/article/pii/S0378779621005186>
179. Du S, Sun Y, Du C. Causality-Based Feature Importance Quantifying Methods: PN-FI, PS-FI and PNS-FI [Internet]. *arXiv [cs.AI]*. 2023. Available from: <http://arxiv.org/abs/2308.14474>
180. Yu K, Guo X, Liu L, Li J, Wang H, Ling Z, et al. Causality-based Feature Selection: Methods and Evaluations. *ACM Comput Surv* [Internet]. 2020 Sep 28;53(5):1–36. Available from: <https://doi.org/10.1145/3409382>
181. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* [Internet]. *IEEE*; 2015. p. 1200–5.

Available from: <http://dx.doi.org/10.1109/MIPRO.2015.7160458>

182. Garner BH, Anderson DJ. Surgical Site Infections: An Update. *Infect Dis Clin North Am* [Internet]. 2016 Dec;30(4):909–29. Available from: <http://dx.doi.org/10.1016/j.idc.2016.07.010>
183. Cheadle WG. Risk factors for surgical site infection. *Surg Infect* [Internet]. 2006;7 Suppl 1:S7–11. Available from: <http://dx.doi.org/10.1089/sur.2006.7.s1-7>
184. Gaynes RP. Surgical-site infections (SSI) and the NNIS Basic SSI Risk Index, part II: room for improvement. *Infect Control Hosp Epidemiol* [Internet]. 2001 May;22(5):266–7. Available from: <http://dx.doi.org/10.1086/501897>
185. Seidelman JL, Mantyh CR, Anderson DJ. Surgical Site Infection Prevention: A Review. *JAMA* [Internet]. 2023 Jan 17;329(3):244–52. Available from: <http://dx.doi.org/10.1001/jama.2022.24075>
186. Lotfinejad N, Peters A, Tartari E, Fankhauser-Rodriguez C, Pires D, Pittet D. Hand hygiene in health care: 20 years of ongoing advances and perspectives. *Lancet Infect Dis* [Internet]. 2021 Aug;21(8):e209–21. Available from: [http://dx.doi.org/10.1016/S1473-3099\(21\)00383-2](http://dx.doi.org/10.1016/S1473-3099(21)00383-2)
187. Anderson BL. Puerperal group A streptococcal infection: beyond Semmelweis. *Obstet Gynecol* [Internet]. 2014 Apr;123(4):874–82. Available from: <http://dx.doi.org/10.1097/AOG.0000000000000175>
188. Wyklicky H, Skopec M. Ignaz Philipp Semmelweis, the prophet of bacteriology. *Infect Control* [Internet]. 1983 Sep-Oct;4(5):367–70. Available from: <http://dx.doi.org/10.1017/s0195941700059762>
189. Ignaz Semmelweis, o médico colocado em manicômio por insistir na importância de se lavar as mãos [Internet]. *G1*. 2019 [cited 2024 Mar 4]. Available from: <https://g1.globo.com/ciencia-e-saude/noticia/2019/09/29/ignaz-semmelweis-o-medico-colocado-em-manicomio-por-insistir-na-importancia-de-se-lavar-as-maos.ghtml>
190. Semmelweis I (1861) [Internet]. The James Lind Library. 2010 [cited 2024 Feb 25]. Available from: <https://www.jameslindlibrary.org/semmelweis-i-1861/>
191. Tyagi U, Barwal KC. Ignac Semmelweis-Father of Hand Hygiene. *Indian J Surg* [Internet]. 2020 Jun;82(3):276–7. Available from: <http://dx.doi.org/10.1007/s12262-020-02386-6>
192. Persson J. Semmelweis's methodology from the modern stand-point: intervention studies and causal ontology. *Stud Hist Philos Biol Biomed Sci* [Internet]. 2009 Sep;40(3):204–9. Available from: <http://dx.doi.org/10.1016/j.shpsc.2009.06.003>
193. Hallett C. The attempt to understand puerperal fever in the eighteenth and early nineteenth centuries: the influence of inflammation theory. *Med Hist* [Internet]. 2005 Jan;49(1):1–28. Available from: <http://dx.doi.org/10.1017/s0025727300000119>
194. Wells J. Puerperal Fever in Britain: Failed Models of Disease Causation [Internet]. University of South Florida; 2010 [cited 2024 Feb 25]. Available from: <https://digitalcommons.usf.edu/etd/3699>
195. Okamura K. Interdisciplinarity revisited: evidence for research impact and dynamism.

- Palgrave Communications [Internet]. 2019 Nov 12 [cited 2024 Feb 25];5(1):1–9. Available from: <https://www.nature.com/articles/s41599-019-0352-4>
196. Sun Y, Latora V. The evolution of knowledge within and across fields in modern physics. *Sci Rep* [Internet]. 2020 Jul 21;10(1):12097. Available from: <http://dx.doi.org/10.1038/s41598-020-68774-w>
 197. Cruz Rivera S, Kyte DG, Aiyegbusi OL, Keeley TJ, Calvert MJ. Assessing the impact of healthcare research: A systematic review of methodological frameworks. *PLoS Med* [Internet]. 2017 Aug;14(8):e1002370. Available from: <http://dx.doi.org/10.1371/journal.pmed.1002370>
 198. Morgan SL, Winship C. *Counterfactuals and Causal Inference* [Internet]. Cambridge University Press; 2015. 499 p. Available from: <https://play.google.com/store/books/details?id=Q6YaBQAAQBAJ>
 199. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* [Internet]. 2016 Apr 15;183(8):758–64. Available from: <http://dx.doi.org/10.1093/aje/kwv254>
 200. Birgand G, Haudebourg T, Grammatico-Guillon L, Moret L, Gouin F, Mauduit N, et al. Intraoperative Door Openings and Surgical Site Infection: A Causal Association? *Clin Infect Dis* [Internet]. 2020 Jul 11;71(2):469–70. Available from: <http://dx.doi.org/10.1093/cid/ciz954>
 201. Seaman SJ, Han E, Arora C, Kim JH. Surgical site infections in gynecology: the latest evidence for prevention and management. *Curr Opin Obstet Gynecol* [Internet]. 2021 Aug 1;33(4):296–304. Available from: <http://dx.doi.org/10.1097/GCO.0000000000000717>
 202. Tanner J, Dumville JC, Norman G, Fortnam M. Surgical hand antisepsis to reduce surgical site infection. *Cochrane Database Syst Rev* [Internet]. 2016 Jan 22;2016(1):CD004288. Available from: <http://dx.doi.org/10.1002/14651858.CD004288.pub3>
 203. Cheng H, Chen BPH, Soleas IM, Ferko NC, Cameron CG, Hinoul P. Prolonged Operative Duration Increases Risk of Surgical Site Infections: A Systematic Review. *Surg Infect* [Internet]. 2017 Aug/Sep;18(6):722–35. Available from: <http://dx.doi.org/10.1089/sur.2017.089>
 204. de Jonge SW, Gans SL, Atema JJ, Solomkin JS, Dellinger PE, Boermeester MA. Timing of preoperative antibiotic prophylaxis in 54,552 patients and the risk of surgical site infection: A systematic review and meta-analysis. *Medicine* [Internet]. 2017 Jul;96(29):e6903. Available from: <http://dx.doi.org/10.1097/MD.0000000000006903>
 205. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package “dagitty.” *Int J Epidemiol* [Internet]. 2016 Dec 1;45(6):1887–94. Available from: <http://dx.doi.org/10.1093/ije/dyw341>
 206. *CausalQueries* [Internet]. [cited 2024 Apr 17]. Available from: <https://integrated-inferences.github.io/CausalQueries/>
 207. *Discovering the* [Internet]. NEC. [cited 2024 Feb 27]. Available from:

<https://www.nec.com/en/global/rd/technologies/202008/index.html>

208. Understanding Causal Research & Why It's Important for Your Business [Internet]. Pollfish Resources. Pollfish; 2021 [cited 2024 Feb 27]. Available from: <https://resources.pollfish.com/market-research/understanding-causal-research-why-its-important-for-your-business/>
209. Hufstedler H, Mauer N, Yeboah E, Carr S, Rahman S, Danzer AM, et al. Application of Causal Inference Methods to Pooled Longitudinal Non- Randomized Studies: A Methodological Systematic Review. *Res Sq* [Internet]. 2023 Aug 30; Available from: <http://dx.doi.org/10.21203/rs.3.rs-3282208/v1>
210. Vandembroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol* [Internet]. 2016 Dec 1;45(6):1776–86. Available from: <http://dx.doi.org/10.1093/ije/dyv341>
211. Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *Eur J Epidemiol* [Internet]. 2021 Sep;36(9):873–87. Available from: <http://dx.doi.org/10.1007/s10654-020-00703-7>
212. Dahabreh IJ, Hernán MA. Extending inferences from a randomized trial to a target population. *Eur J Epidemiol* [Internet]. 2019 Aug;34(8):719–22. Available from: <http://dx.doi.org/10.1007/s10654-019-00533-2>
213. Zhang M, Wang Y. Adjusting for observational secondary treatments in estimating the effects of randomized treatments. *Biostatistics* [Internet]. 2013 Jul;14(3):491–501. Available from: <http://dx.doi.org/10.1093/biostatistics/kxs060>
214. Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials. Part II: an application in a clinical trial of unresectable non-small-cell lung cancer. *Stat Med* [Internet]. 2004 Jul 15;23(13):2005–22. Available from: <http://dx.doi.org/10.1002/sim.1817>
215. Zheng C, Dai R, Gale RP, Zhang MJ. Causal inference in randomized clinical trials. *Bone Marrow Transplant* [Internet]. 2020 Jan;55(1):4–8. Available from: <http://dx.doi.org/10.1038/s41409-018-0424-x>
216. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* [Internet]. 2018 Aug;210:2–21. Available from: <http://dx.doi.org/10.1016/j.socscimed.2017.12.005>
217. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA* [Internet]. 2022 Dec 27;328(24):2446–7. Available from: <http://dx.doi.org/10.1001/jama.2022.21383>
218. Vandembroucke J, Pearce N. Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? *Am J Epidemiol* [Internet]. 2015 Nov 15;182(10):826–33. Available from: <http://dx.doi.org/10.1093/aje/kwv225>
219. Gianicolo EAL, Eichler M, Muensterer O, Strauch K, Blettner M. Methods for Evaluating Causality in Observational Studies. *Dtsch Arztebl Int* [Internet]. 2020 Feb 14;116(7):101–7. Available from: <http://dx.doi.org/10.3238/arztebl.2020.0101>

220. Coscia Requena C, Muriel A, Peñuelas O. Analysis of causality from observational studies and its application in clinical research in Intensive Care Medicine. *Med Intensiva* [Internet]. 2018 Jun;42(5):292–300. Available from: <https://www.medintensiva.org/en-analysis-causality-from-observational-studies-articulo-S2173572718300857>
221. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology* [Internet]. 2001 May;12(3):313–20. Available from: <http://dx.doi.org/10.1097/00001648-200105000-00011>
222. Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol* [Internet]. 2002 Feb;31(1):163–5. Available from: <http://dx.doi.org/10.1093/ije/31.1.163>
223. Etminan M, Collins GS, Mansournia MA. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. *Chest* [Internet]. 2020 Jul;158(1S):S21–8. Available from: <http://dx.doi.org/10.1016/j.chest.2020.03.011>
224. Tsuda T, Babazono A, Shigemi J, Otsu T, Mino Y. [Causal inference in medicine--decision making]. *Sangyo Eiseigaku Zasshi* [Internet]. 2001 Sep;43(5):161–73. Available from: <http://dx.doi.org/10.1539/sangyoeisei.kj00001991709>
225. Kramer MS, Lane DA. Causal propositions in clinical research and practice. *J Clin Epidemiol* [Internet]. 1992 Jun;45(6):639–49. Available from: [http://dx.doi.org/10.1016/0895-4356\(92\)90136-b](http://dx.doi.org/10.1016/0895-4356(92)90136-b)
226. Olarte Parra C, Bertizzolo L, Schroter S, Dechartres A, Goetghebeur E. Consistency of causal claims in observational studies: a review of papers published in a general medical journal. *BMJ Open* [Internet]. 2021 May 20;11(5):e043339. Available from: <http://dx.doi.org/10.1136/bmjopen-2020-043339>