

Epidemiologia aplicada básica

● Edison Capp
Otto Henrique Nienov
Organizadores

Caroline Darski
Cristiana Palma Kuhl
Fernanda Dapper Machado
Fernanda Santos Grossi
Fernanda Vargas Ferreira
Hellen Meiry Grosskopf Werka
Johanna Ovalle Diaz
Laís Lima Ferreira
Lilian de Sá Paz Ramos
Luciana Stein e Silva
Marina Petter Rodrigues
Michele Strelow Moreira
Nadine de Souza Ziegler
Pedro Henrique Comerlato
Sinara Santos
Vânia Marisia Santos Fortes dos Reis

Universidade Federal do Rio Grande do Sul
Faculdade de Medicina
Programa de Pós-Graduação em Ciências da Saúde:
Ginecologia e Obstetrícia

Epidemiologia Aplicada Básica

Porto Alegre 2021
UFRGS

U58e Universidade Federal do Rio Grande do Sul. Faculdade de Medicina. Programa de Pós-Graduação em Ciências da Saúde: Ginecologia e Obstetrícia Epidemiologia aplicada básica/ Universidade Federal do Rio Grande do Sul; organizadores: Edison Capp e Otto Henrique Nienov – Porto Alegre: UFRGS, 2021.

273p.

ISBN: 978-65-86232-63-9

E-Book: 978-65-86232-64-6

1. Epidemiologia e Bioestatística 2. Estatística 3. Métodos epidemiológicos I. Capp, Edison, org. II. Nienov, Otto Henrique, org. III Título.

NLM: WA950

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
(Bibliotecária Shirlei Galarça Salort – CRB10/1929)

Endereço:

PPG em Ciências da Saúde: Ginecologia e Obstetrícia

FAMED – UFRGS

Rua Ramiro Barcellos, 2400/2º andar

CEP 900035-003 – Porto Alegre – RS

Telefone: +55 51 3308 5607

E-mail: ppggo@ufrgs.br

Editoração: Edison Capp e Otto Henrique Nienov

Capa e diagramação: Edison Capp

Imagem da capa: www.freepik.com/starline

Edison Capp
Otto Henrique Nienov
Organizadores

Caroline Darski
Cristiana Palma Kuhl
Fernanda Dapper Machado
Fernanda Santos Grossi
Fernanda Vargas Ferreira
Hellen Meiry Grosskopf Werka
Johanna Ovalle Diaz
Laís Lima Ferreira
Lilian de Sá Paz Ramos
Luciana Stein e Silva
Marina Petter Rodrigues
Michele Strelow Moreira
Nadine de Souza Ziegler
Pedro Henrique Comerlato
Sinara Santos
Vânia Marisia Santos Fortes dos Reis

Universidade Federal do Rio Grande do Sul

Reitor: Prof. Carlos André Bulhões

Vice-Reitora: Profa. Patrícia Helena Lucas Pranke

Faculdade de Medicina

Diretora: Profa. Lúcia Maria Kliemann

Vice-Diretor: Prof. Luciano Zubaran Goldani

Departamento de Ginecologia e Obstetrícia

Chefe: Profa. Solange Garcia Accetta

Chefe Substituta: Profa. Suzana Arenhart Pessini

PPG em Ciências da Saúde: Ginecologia e Obstetrícia

Coordenador: Prof. Edison Capp

Coordenador substituto: Prof. Eduardo Pandolfi Passos

Hospital de Clínicas de Porto Alegre

Diretora Presidente: Profa. Nadine Clausell

Diretor Médico: Prof. Milton Berger

Diretor Administrativo: Jorge Bajerski

Coordenador GENS: Prof. José Geraldo Lopes Ramos

Coordenadora GPPG: Profa. Patrícia Ashton Prolla

Serviço de Ginecologia e Obstetrícia

Chefe: Profa. Maria Celeste Osório Wender

Serviço de Mastologia

Chefe: Prof. Jorge Villanova Biazús

Centro de Estudos do Serviço de Ginecologia e Obstetrícia

Presidente: Prof. Jorge Villanova Biazús

Fundação Médica do Rio Grande do Sul

Presidente: Prof. Fernando Gomes

*“Não há fatos eternos, como não
há verdades absolutas”.*

Friedrich Nietzsche
1844-1900

Autores

Caroline Darski

Graduação em Fisioterapia, Universidade Federal do Rio Grande do Sul (UFRGS). Mestrado no Programa de Pós-Graduação em Ciências da Saúde Ginecologia e Obstetrícia (PPGGO), UFRGS. Doutoranda PPGGO, UFRGS.

Cristiana Palma Kuhl

Graduação em Fisioterapia, Centro Universitário Metodista IPA. Mestrado no Programa de Pós-Graduação em Medicina: Ciências Cirúrgicas, UFRGS. Doutoranda PPGGO, UFRGS

Edison Capp

Graduação em Medicina, UFRGS. Residência Médica em Ginecologia e Obstetrícia, Hospital Materno-Infantil Presidente Vargas, Porto Alegre. Doutor em Medicina, Universidade Ludwig-Maximilian, Munique, Alemanha. Professor do Departamento de Ginecologia e Obstetrícia e do PPGGO, FAMED, UFRGS.

Fernanda Dapper Machado

Graduação em Farmácia, Universidade Feevale. Mestrado PPGGO. Doutoranda PPGGO, UFRGS.

Fernanda Santos Grossi

Graduação em Medicina pela Universidade Federal de Ciências da Saúde de Porto Alegre, UFCSPA. Residência Médica em Ginecologia e Obstetrícia no Complexo Hospitalar Santa Casa de Porto Alegre. Mestrado PPGGO, UFRGS. Doutoranda PPGGO, UFRGS. Coordenadora do Serviço de Ginecologia e Obstetrícia e preceptora da Residência Médica de Ginecologia e Obstetrícia do Hospital Geral de Caxias do Sul.

Fernanda Vargas Ferreira

Graduação em Fisioterapia, Universidade Federal de Santa Maria (UFSM). Doutorado e Pós-Doutorado PPGGO, UFRGS. Professora do Curso de Fisioterapia, Universidade Federal do Pampa (UNIPAMPA).

Hellen Meiry Grosskopf Werka

Graduação em Medicina, Universidade Federal de Pelotas (UFPel). Residente de Patologia Cirúrgica, Hospital de Clínicas de Porto Alegre, HCPA. Mestranda, PPGGO, UFRGS.

Johanna Ovalle Diaz

Graduação em Medicina, Universidade Militar Nueva Granada, Bogotá, Colombia, UMNG. Graduação em Urologia, Fundação Universitaria Sanitas, Bogotá, Colombia. Pos-Graduação em Urologia Infantil e Reconstructiva, HCPA. Mestranda PPPGO, UFRGS.

Laís Lima Ferreira

Graduação em Nutrição, UFRGS, Mestrado em Endocrinologia, UFRGS. Doutoranda PPGGO, UFRGS.

Lilian de Sá Paz Ramos

Graduação em Medicina, Universidade Federal da Bahia (UFBA). Residência Médica em Ginecologia e Obstetrícia, Hospital Universitário Prof. Edgard Santos, UFBA. Especialização em Mastologia, Hospital Aristides Maltez, Bahia, e pela Sociedade Brasileira de Mastologia. Mestranda PPGGO, UFRGS.

Luciana Stein e Silva

Graduação em Farmácia, Universidade Luterana do Brasil, ULBRA. Pós-graduação em Farmácia Hospitalar, Instituto de Educação e Pesquisa do Hospital Moinhos de Vento, HMV. Mestranda PPGGO, UFRGS.

Marina Petter Rodrigues

Graduação em Fisioterapia, UFRGS. Especialização em Fisioterapia Pélvica, Faculdade Inspirar, PoA. Mestrado e Doutoranda PPGGO, UFRGS.

Michele Strelow Moreira

Graduação em Licenciatura Plena em Matemática, Universidade Católica de Pelotas. Graduação em Farmácia, UFPel. Pós-Graduanda em Farmácia Hospitalar e Clínica, Faculdade Ciências da Saúde HMV. Mestranda PPGGO, UFRGS.

Nadine de Souza Ziegler

Graduação em Medicina, UFSM. Especialista em Ginecologia e Obstetrícia e Medicina Fetal pela Febrasgo. Especialista em Ultrassom em Ginecologia e Obstetrícia pelo Conselho Brasileiro de Radiologia. Mestranda PPGGO, UFRGS.

Otto Henrique Nienov

Graduação em Toxicologia Analítica, UFCSPA. Graduação em Biomedicina, Universidade FEEVALE. Mestrado, Doutorado e Pós-Doutorado PPGGO, UFRGS.

Pedro Henrique Comerlato

Graduação em Medicina, UFCSPA. Residência Médica em Medicina Interna e em Medicina Intensiva, HCPA/UFRGS. Doutorando PPG Endocrinologia, UFRGS.

Sinara Santos

Graduação em Psicologia, ULBRA. Especialista em Psicologia Hospitalar, HCPA. Mestrado PPGGO. Doutoranda PPGGO, UFRGS.

Vânia Marisia Santos Fortes dos Reis

Graduação em Biomedicina, Mestrado em Ciências Biológicas: Fisiologia, Doutoranda PPGGO, UFRGS. Aluna de graduação em Medicina, UFCSPA.

Sumário

Apresentação	15
Aplicando a epidemiologia	19
O que é epidemiologia?	19
Bases de dados: atualizando o conhecimento	20
A pergunta de pesquisa: o ponto de partida	21
Estatística: um fator essencial	22
Classificação dos desenhos de estudo	28
Quanto à originalidade	28
Quanto ao tipo de unidade de estudo	28
Quanto à inferência	29
Quanto ao perfil de avaliação	30
Quanto ao período de seguimento	30
Quanto à direção do tempo	31
Quanto à existência de controle ou comparador	31
Quanto aos tipos de frequência	32
Quanto à procedência da equipe	34
Estudo piloto	34
Hierarquia dos níveis de evidência	34
Considerações éticas	36
Análise crítica e qualidade metodológica de artigos científicos	39
O que é pesquisa científica?	39
Medicina Baseada em Evidências	41
Por que ler um artigo científico?	41
Estrutura de um artigo científico	44
Como escrever um artigo?	46
Dúvida: ler ou não um artigo	47
Avaliando a qualidade metodológica	49
O estudo apresenta originalidade?	49
O objetivo do estudo é claro e apropriado?	49
O estudo é sobre quem?	50
O estudo foi bem delineado?	51
O viés sistemático foi evitado?	52
O estudo teve tamanho e tempo adequados?	58
Estudos secundários: revisão sistemática e metanálise	63
Quando uma revisão é sistemática?	64
Revisão sistemática	66
Por que fazer uma revisão sistemática?	67

Quanto tempo leva uma revisão sistemática?	68
Vantagens da revisão sistemática	69
Limitações da revisão sistemática.....	70
Como conduzir e avaliar uma revisão sistemática?	70
Etapa 1: Formulação da pergunta de pesquisa.....	71
Definição da pergunta de pesquisa.....	71
Protocolo da revisão sistemática.....	74
Guias metodológicos.....	75
A revisão sistemática necessita de recursos para a pesquisa?.....	77
Etapa 2: Estratégia de busca na literatura.....	78
Escolha das bases de dados.....	79
Elaboração da estratégia de busca.....	80
Literatura cinzenta.....	84
Atualização da busca.....	85
Etapa 3: Seleção dos estudos.....	85
Gerenciadores de referências.....	85
Triagem dos estudos.....	87
Etapa 4: Extração dos dados.....	90
Estudos versus publicações.....	90
Definindo as variáveis.....	90
Dados ausentes: e agora?.....	91
Registro da extração de dados.....	92
Ciência aberta.....	93
Etapa 5: Avaliação da qualidade metodológica dos estudos.....	95
Limitação metodológica.....	95
Qualidade metodológica.....	96
Guias para avaliação da qualidade dos estudos.....	97
Qualidade metodológica versus qualidade da evidência.....	101
Etapa 6: síntese dos dados (metanálise).....	102
Definição de metanálise.....	102
Como ler um gráfico de floresta (forest plot)?.....	104
Dados individuais dos participantes.....	107
Metanálise direta.....	107
Inverso da variância.....	107
Efeitos fixos e aleatórios.....	108
Metanálise em rede.....	109
Outras abordagens de análise.....	110
Softwares para metanálise.....	110
Heterogeneidade.....	111
Análise de subgrupo.....	113
Metarregressão.....	113
Viés de publicação.....	114

Exclusão de outliers	116
Etapa 7: avaliação da qualidade das evidências	116
Sistema GRADE.....	117
Classificação dos desfechos.....	117
Domínios do sistema GRADE	118
Níveis de evidência	118
Apresentação dos dados	122
Força de recomendação	122
Avaliação de evidências qualitativas.....	123
Etapa 8: redação dos resultados	124
Etapa 9: publicação e atualização dos resultados	126
Estudos secundários: diretriz clínica e análise econômica	129
Diretrizes Clínicas.....	129
Mas, por que diretriz clínica?	131
Evento adverso	132
Por que seguir as diretrizes clínicas?.....	132
Desenvolvimento de diretrizes clínicas	132
Avaliação da qualidade da diretriz clínica.....	133
Análise econômica em saúde	136
Análise de custo-minimização.....	136
Análise de custo-efetividade.....	136
Análise de custo-utilidade.....	137
Análise de custo-benefício.....	138
Análise de impacto orçamentário	138
Elaboração de análise econômica.....	138
Avaliação da qualidade da análise econômica	139
Estudos de intervenção: ensaios de campo, comunitário e clínico	143
Ensaio de campo	143
Ensaio comunitário	144
Ensaio clínico	144
Fases do ensaio clínico	145
Ensaio clínico randomizado	146
Por que o ECR é considerado padrão-ouro?.....	148
Etapas de um ECR.....	148
Registro do protocolo clínico.....	149
Tamanho de amostra de um ECR.....	149
Critérios de elegibilidade	150
Braços do estudo.....	150
Randomização ou aleatorização	151
Placebo e o efeito placebo.....	154

Implicações éticas do uso de placebo	155
Estudos de não-inferioridade, de equivalência e de superioridade	155
Mascaramento ou cegamento.....	156
Implicações éticas do cegamento	157
Desfechos primários e substitutos.....	158
Delineamentos de ECR.....	159
Ensaio clínico controlado aleatorizado ou paralelo	159
Ensaio clínico controlado pareado.....	159
Ensaio clínico controlado sequencial	159
Ensaio clínico controlado cruzado	159
Ensaio clínico controlado fatorial	160
Quando o ECR deve ser considerado?.....	160
Quando o ECR não deve ser realizado?.....	160
Vantagens e desvantagens do ECR.....	161
Quando um ensaio clínico não é randomizado	162
Como redigir um ECR?	163
Como avaliar a qualidade metodológica de um ECR?.....	163
Avaliação crítica de um ECR	164
Estudos de testes diagnósticos ou de rastreamento.....	171
Testes de rastreamento.....	171
Testes diagnósticos.....	172
Teoria da probabilidade.....	173
Acurácia, precisão e generalização.....	173
Validação de testes diagnósticos ou de rastreamento	174
Tabela de contingência 2 X 2	175
Sensibilidade e especificidade	176
Curva ROC (<i>Relative Operating Characteristics</i>).....	178
Acurácia	180
Prevalência.....	180
Probabilidade pré-teste	180
Valores preditivos positivo e negativo.....	180
Razões de verossimilhanças positivo negativo	181
Qual parâmetro é mais útil para avaliar um teste índice?	182
Qualidade de estudos de testes diagnósticos/rastreamento.....	183
Avaliando estudos de testes diagnósticos e rastreamento	184
Validade do estudo.....	184
Precisão do teste índice.....	187
Aplicabilidade do teste índice.....	189

Estudos observacionais: coorte, caso-controle e transversal	191
Estudos de coorte.....	191
Modelo conceitual do estudo de coorte	192
Fases do estudo de coorte	194
Como fazer a seleção do grupo de comparação?	194
Amostragem	195
O que acontece com alguém que tem a doença?	195
Exposição, desfecho e fatores de confusão	197
Medição dos dados	198
Temporalidade nos estudos de coorte	199
Duração de um estudo de coorte.....	200
Quando os estudos de coorte são indicados?	200
Incidência, risco relativo e risco atribuível	201
Vantagens e desvantagens dos estudos de coorte	203
Vieses em estudos de coorte	204
Redação de estudos de coorte.....	205
Análise crítica de estudos de coorte	206
Análise crítica de estudos de prognóstico.....	213
Estudo de caso-controle	215
Modelo conceitual do estudo de caso-controle.....	215
Fases do estudo caso-controle	217
Seleção de casos e controles.....	217
Caso-controle de base populacional aninhado	219
Pareamento ou emparelhamento dos grupos.....	220
Amostragem	221
Medida de associação: <i>Odds Ratio</i>	222
Vantagens e desvantagens dos estudos de caso-controle.....	223
Vieses dos estudos de casos-controle.....	224
Redação dos estudos de caso-controle.....	225
Análise crítica dos estudos de caso-controle.....	225
Estudo transversal.....	227
O que é prevalência?.....	229
População de estudo.....	229
Tamanho da amostra	229
Fonte de dados	229
Fatores de exposição e desfecho	230
Modelo conceitual do estudo transversal.....	230
Fases do estudo transversal.....	231
Razão de prevalência.....	231
Quais são os critérios para julgar causalidade?.....	234

Vantagens e desvantagens do estudo transversal.....	234
Vieses do estudo transversal	235
Redação de estudo transversal.....	235
Análise crítica do estudo transversal	235
Estudo Ecológico.....	236
Estudos descritivos (relato e série de casos) e estudos pré-clínicos (<i>in silico</i>, <i>in vitro</i> e <i>in vivo</i>).....	239
Relato de caso e série de casos.....	239
Qual a diferença entre relato de caso e série de casos?	240
Indicação para elaboração do relato de caso e da série de casos	240
Estrutura do relato de caso e da série de casos.....	241
Vantagens e desvantagens do relato de caso e da série de casos	243
Redação do relato de caso e da série de casos	244
Análise crítica do relato de caso e da série de casos	244
Estudos pré-clínicos.....	246
Estudos <i>in silico</i>	247
Estudos <i>in vitro</i>	248
Estudos <i>ex vivo</i>	249
Estudos <i>in vivo</i>	249
Redação de estudos pré-clínicos	252
Análise crítica dos estudos pré-clínicos	252
Pesquisa qualitativa	255
O que é pesquisa qualitativa?	256
Definição da pergunta de pesquisa.....	256
Formulação da hipótese	257
Objeto de estudo	258
Elaboração dos objetivos	258
Existe amostra na pesquisa qualitativa?	259
Etnografia	259
Estudo <i>in loco</i>	260
Universo empírico.....	261
Métodos de coleta de dados	261
Validade da pesquisa qualitativa	264
E como saber se a pesquisa qualitativa atingiu a saturação teórica?	265
Análise de dados	268
Vantagens e desvantagens da pesquisa qualitativa	268
Redação da pesquisa qualitativa	269
Análise crítica de pesquisa qualitativa.....	269

Apresentação

O Programa de Pós-Graduação em Ciências da Saúde: Ginecologia e Obstetrícia (PPGGO) completou recentemente 5 anos de existência e, esta obra, é mais uma demonstração da potencialidade em produzir informação científica e de formar recursos humanos de alta qualidade. Parabéns a todos que participaram da criação e da trajetória do PPGGO!

Este livro é complementar ao livro “Bioestatística Quantitativa Aplicada” e visa colaborar na formação do pós-graduando no ensino e na pesquisa, principalmente no que tange a análise crítica de artigos científicos dos diversos delineamentos de estudos abordados neste livro. Despretensioso e reflexivo, o livro originou-se a partir da disciplina “Epidemiologia Básica Aplicada às Ciências da Saúde”, ministrada pelo Dr. Otto Nienov e com a participação dos demais autores atuando como monitores. Foi assistida por mais de 150 alunos de pelo menos 21 diferentes programas de pós-graduação da UFRGS e de outras universidades. As aulas atravessaram os limites estaduais, tendo participantes do Sul ao Norte do Brasil, e nacionais, com alunos assistindo as aulas remotamente da Argentina e do Canadá. Foi toda ministrada de maneira virtual e híbrida e, apesar de tantos participantes, funcionou extraordinariamente bem. Houve oportunidade de esclarecimento de dúvidas, partilha de conhecimentos entre os atores envolvidos e, mesmo sendo a primeira experiência de alguns com atividades remotas, as participações foram feitas com etiqueta que deve servir de exemplo para outras atividades neste formato.

Agradecimentos a todos que participaram da realização desta atividade.

Edison Capp
Otto Henrique Nienov
Organizadores

Apresentação FAMED

Em um trabalho de absoluta relevância, o Programa de Pós-Graduação em Ciências da Saúde: Ginecologia e Obstetrícia, o nosso PPGGO, nos oferece mais uma bela obra que em muito pode auxiliar nossos alunos e qualificar o nosso pós-graduação. O “Epidemiologia Aplicada Básica” é fruto de um trabalho primoroso que estimula a integração entre os alunos de pós-graduação, desenvolve a habilidade da escrita científica e oferece um produto muito rico, a ser utilizado em diversos níveis da formação em pesquisa científica embasada, apresentando desde a justificativa histórica até a aplicação final das mais importantes ferramentas de análise crítica, indispensáveis a quem pretende fazer ciência de qualidade.

Vivendo em uma realidade de pandemia, pudemos experimentar a importância da informação precisa, baseada em dados confiáveis, para planejamento de políticas públicas, quer seja do ponto de vista de práticas de saúde, quanto do ponto de vista sócio-econômico.

A partir de uma experiência muito bem-sucedida com a elaboração do livro de Estatística, esta obra complementar nos mostra que, mesmo em períodos muito novos, em que todos passamos a experimentar o ensino remoto e a nos adaptar a novas ferramentas de ensino e comunicação, é possível pensar, inovar, produzir e colaborar de diversas formas com redes de geração de conhecimento, indispensáveis à evolução de nossa sociedade.

Parabéns PPGGO!

Obrigada PPGGO!

Lúcia Maria Kliemann
Diretora FAMED/UFRGS

Aplicando a epidemiologia

*Fernanda Santos Grossi
Marina Petter Rodrigues
Edison Capp
Otto Henrique Nienov*

O que é epidemiologia?

A palavra “epidemiologia” tem origem nas palavras gregas: *epi* “sobre”, *demio* “povo” e *logos* “estudo” (Figura 1). A epidemiologia pode ser definida como o estudo de diferentes fatores que interferem na ocorrência e distribuição de eventos relacionados à saúde em populações específicas, propondo intervenções que busquem a prevenção e o controle de doenças e disfunções. Embora a epidemiologia se preocupe com o curso, desfechos e impacto das doenças, é importante entender que ela não se baseia apenas nisso, mas também na saúde e qualidade de vida, ou seja, ela engloba o estudo do processo de saúde-doença, produzindo o conhecimento necessário para promover a saúde individual através de medidas de alcance coletivo.

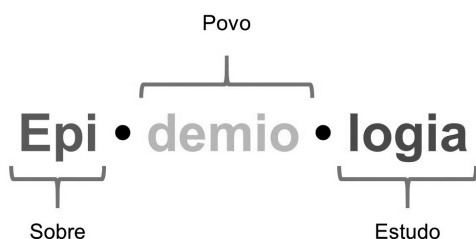


Figura 1. Significado da palavra “epidemiologia”.

Medir saúde e doença é fundamental para a aplicação prática da epidemiologia. Segundo a Organização Mundial da Saúde, a “saúde é um estado de completo bem-estar físico, mental e social e não apenas a mera ausência de doença”. Existem

diversas medidas para se caracterizar a saúde das populações, como critérios diagnósticos, medidas de ocorrência (incidência e prevalência), medidas de risco, entre outras.

Bases de dados: atualizando o conhecimento

Para que o conhecimento seja suficiente para o desenvolvimento de estratégias coerentes e eficazes, é necessário tempo e investimento constante em uma gama de pesquisas que sigam rigorosas metodologias e delineamentos, abrangendo diversas áreas que não apenas a das ciências da saúde. Ainda que muito já se saiba sobre diversas condições, é permanente a necessidade de se gerar novas informações a todo momento.

Considerando esta incessante necessidade de atualização do conhecimento somada à velocidade com que isto circula, nos deparamos com uma situação: o excesso de informação. Em um primeiro momento, isso parece muito positivo, mas precisamos ter um olhar atento para este acúmulo de evidências. O acesso à informação aumentou consideravelmente nos últimos anos, principalmente com uso da internet como fonte de informação. Mas, a sofisticação da busca e a eficiência ao encontrar respostas não cresceram com a mesma eficiência.

Diversas são as bases de dados que temos acesso para a busca de artigos científicos na área da saúde, como *PubMed*, *Lilacs*, *Scielo*, *Web of Science*, *Cochrane*, *Scopus*, entre outros. Além disso, existem plataformas que se assemelham mais a livros-texto ou manuais clínicos eletrônicos detalhados, baseados em evidências, atualizados continuamente e projetados para serem acessíveis aos usuários. Dentre essas plataformas, podemos citar o *Clinical Evidence*, *DynaMed* e *American College of Physicians Physicians' Information and Education Resource* (ACP PIER). Alguns destes, como o PIER e o *DynaMed*, possuem aplicativos que facilitam o uso em assistentes pessoais digitais ou outros equipamentos portáteis manuais, o que melhora q sua utilidade.

Muitas das bases de dados são práticas, facilmente acessíveis e contêm artigos de conceituados periódicos. Os "livros-texto" citados aspiram ser firmemente baseados em evidências, revisados por pares, com regularidade e com links

para pesquisa primária incorporada às suas recomendações. Pensando assim, será que isso é suficiente para confiarmos em uma evidência científica? Não necessariamente. A proliferação de bases de dados com milhares de evidências disponíveis faz com que o leitor esteja em meio a uma selva de informações, sendo extremamente importante que o pesquisador tenha uma boa base de conhecimento metodológico, a fim de “filtrar” as suas leituras e analisar criticamente os resultados e conclusões.

Sendo assim, a epidemiologia aplicada às ciências da saúde é o estudo das metodologias de pesquisa e serve como guia importante na hora de ler um artigo científico. Reconhecer os diferentes tipos de delineamento, seus aspectos metodológicos e estatísticos e aplicar estes conceitos na leitura crítica e rigorosa dos manuscritos é essencial não só para quem faz pesquisa, mas também para quem atua na parte clínica do processo de saúde-doença, pois a famosa “medicina baseada em evidências” que tanto se ouve falar, nada mais é do que aliar as evidências científicas ao conhecimento clínico e às preferências do paciente.

A pergunta de pesquisa: o ponto de partida

Tanto para o desenvolvimento de pesquisas científicas como para a condução de decisões clínicas baseadas em evidências, o primeiro passo é a formulação da pergunta de pesquisa, ou seja, o que se quer saber? Este é o ponto de partida. De uma ideia surge um problema, que é traduzido em uma pergunta de pesquisa. A pergunta da pesquisa é transformada em um objetivo e, o pesquisador ou investigador, formula uma hipótese para responder à sua pergunta e resolver o problema da pesquisa (Figura 2).

Para uma maior solidez da evidência, tanto no desenvolvimento de um projeto de pesquisa como na escrita ou leitura de um artigo científico, é necessário que nos atentemos a alguns pilares que embasam o rigor metodológico. Guie e direcione a sua investigação, mas primeiro pense: a pergunta é factível de ser respondida? Foque no seu objetivo. A pergunta, além de relevante, deve ser algo em que você realmente esteja interessado não apenas por curiosidade, mas sim, porque você

conhece sobre o problema e identifica onde está a fronteira entre o conhecimento atual e o desconhecimento. Dominar a literatura publicada em uma área de estudo é necessário para uma boa investigação. Assim, o primeiro passo para a identificação da pergunta é estar familiarizado suficientemente com um assunto. Isso fará com que você consiga delinear o seu estudo com mais facilidade e coerência.

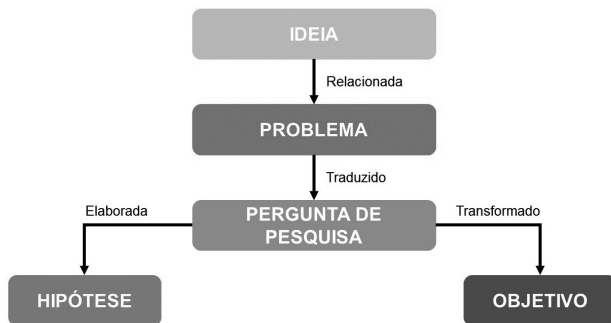


Figura 2. Processo do desenvolvimento da pesquisa científica.

Com estes pilares bem esclarecidos, partimos com maior segurança para a formulação de um projeto de pesquisa e também para a leitura de um artigo científico de forma crítica. Fazer a pergunta de forma equivocada, buscar respostas nas fontes erradas ou não saber analisá-las seguindo critérios metodológicos, é semelhante a não ter lido artigo algum.

Estatística: um fator essencial

Comumente, a estatística é apontada como o aspecto de mais difícil entendimento. A principal questão é: como quem não é estatístico avalia os testes estatísticos? Não pule a análise estatística e nem delegue essa função para outra pessoa. Entenda que ela é um fator essencial para a confirmação e veracidade dos achados, evitando avaliações tendenciosas. Ao realizar a leitura de um artigo científico, não se pode deixar os aspectos estatísticos de lado e se apoiar simplesmente na descrição dos resultados pelos autores (pois sim, eles podem tanto analisar como interpretar os resultados de forma inadequada). Por isso, avalie criteriosamente as tabelas, gráficos e testes estatísticos utilizados.

O primeiro aspecto estatístico a se avaliar em um artigo científico é: os autores fizeram um cálculo do tamanho de amostra? Determinar o tamanho da amostra, que é um subconjunto da população, é crucial para o desenvolvimento de um estudo, possibilitando maior precisão nos resultados e fornecendo embasamento adequado à tomada de decisões.

Quais tipos de dados foram obtidos? Foram utilizados os testes estatísticos adequados? É importante entender e classificar as variáveis de estudo para analisar e descrever os resultados e determinar os testes estatísticos a serem executados (Figura 3).

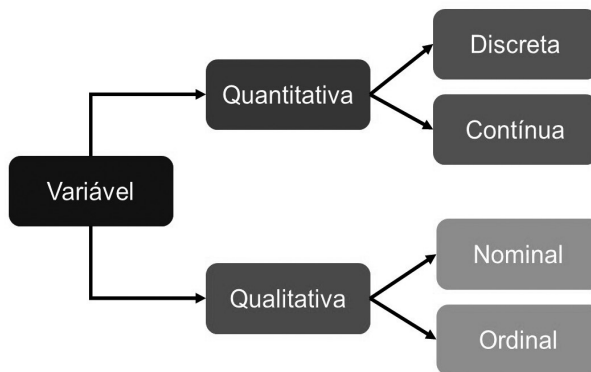


Figura 3. Tipos de variáveis.

A figura 4 resume e é bastante útil para se verificar quais testes estatísticos são adequados conforme o tipo de amostra, número de grupos comparados e distribuição dos dados. Lembre-se que a distribuição de normalidade dos dados deve ser avaliada pelo teste de Shapiro-Wilk, que é considerado como de melhor potência. Se a significância do teste for superior a 0,05 ($p > 0,05$), as variáveis seguem uma distribuição normal e utiliza-se testes paramétricos; se a significância do teste for inferior ou igual a 0,05 ($p \leq 0,05$), as variáveis seguem uma distribuição não normal e utiliza-se testes não paramétricos.

A escolha adequada do teste (paramétrico ou não paramétrico) a ser utilizado é de fundamental importância, pois conclusões erradas podem ser obtidas se optarmos pelo teste

inapropriado. Além disso, a apresentação dos dados deve estar de acordo com a suposição de distribuição da variável: se distribuição normal, utilizamos média e desvio padrão; se distribuição não normal, expressamos os dados em mediana e quartis ou valores mínimo e máximo.

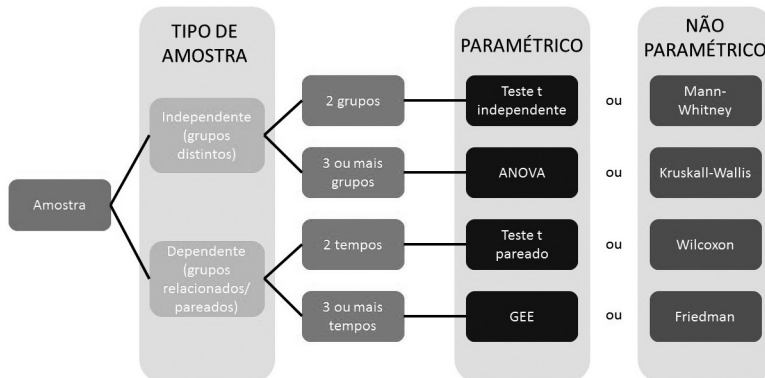


Figura 4. Testes paramétricos e não paramétricos para amostras independentes e pareadas (Mancuso e cols., 2017).

Não se deixe enganar por testes estatísticos que você não conheça. Se o artigo descreve um conjunto-padrão de dados coletados de modo padronizado, mas o teste utilizado é impronunciável ou a forma como os dados foram analisados parece obscura, desconfie. Os autores devem justificar e fornecer uma referência para o teste usado. Além disso, é importante verificar se os dados foram analisados de acordo com o protocolo original do estudo que, por exemplo, nos ensaios clínicos randomizados, pode estar registrado no ClinicalTrials (<https://clinicaltrials.gov/>).

Ao avaliar a forma como os dados foram apresentados, verifique se estes foram analisados de acordo com o objetivo inicial do estudo. É preciso ter muito cuidado com a prática comum dos pesquisadores de “mineração dos dados”. Análises subestimadas ou superestimadas podem levar a uma inflação dos resultados e, conseqüentemente, conclusões erradas. Lembre-se: se torturarmos inapropriadamente os números, eles confessam qualquer coisa.

Verifique se foram cumpridos os parâmetros para o teste de hipóteses. As hipóteses estatísticas são suposições ou afirmações feitas sobre um parâmetro, ou parâmetros, na população. Em qualquer teste de hipóteses existem duas hipóteses: a hipótese nula ou de nulidade (H_0) e a hipótese alternativa (H_A). A H_0 estabelece uma ausência de diferença entre os parâmetros avaliados, enquanto que a H_A corresponde a hipótese que o pesquisador quer confirmar e é contrária à H_0 .

O teste de hipóteses pode ser unilateral à esquerda ou à direita ou bicaudal, sendo que o termo “cauda” se refere aos extremos da distribuição. De modo geral, a análise estatística deve testar a hipótese de que tanto valores altos como baixos em seu conjunto de dados sejam devidos ao acaso, o que significa que um teste bicaudal deve ser aplicado, a menos que existam evidências de que a diferença somente pode ocorrer em uma direção (Figura 5).

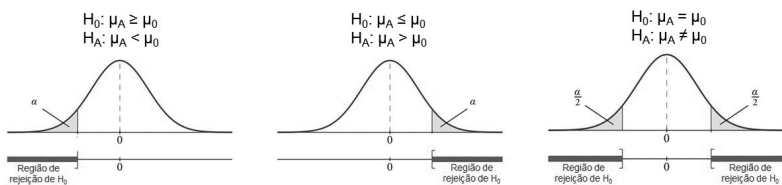


Figura 5. Teste de hipóteses, em que μ_A refere-se a média da população amostrada, μ_0 a média da população referência e α o nível de significância

Os valores extremos, também conhecidos como valores atípicos ou *outliers*, foram analisados com bom senso e com ajustes estatísticos apropriados? Valores atípicos distorcem a média e inflacionam o desvio padrão, sendo necessária uma análise criteriosa dos dados para detectá-los. Entender os *outliers* é fundamental em uma análise de dados.

A correlação foi diferenciada da regressão? O coeficiente de correlação foi calculado e interpretado corretamente? Para muitos, os termos correlação e regressão são sinônimos. Porém, tanto a regressão como a correlação são termos estatísticos precisos, que servem para funções bastante diferentes. A correlação nos indica a direção e intensidade de uma associação. Não implica causalidade, por mais forte que seja a correlação. A regressão

nos informa o impacto de uma variável (simples) ou conjunto de variáveis (múltipla) sobre um desfecho, ou seja, permite avaliar se uma variável dependente é predita a partir de outra(s) variável(is) independente(s). Verifique também se o modelo de regressão foi adequado para o delineamento do estudo e tipo de variável dependente (se quantitativa e dicotômica) e, se realmente era necessário ou possível de ser realizado. Da mesma forma, nas correlações, examine se o tipo de correlação foi apropriado para a distribuição dos dados (se normal ou não) e tipos de variáveis (quantitativa, ordinal ou nominal).

Foram feitas suposições sobre a natureza e a direção da causalidade? A correlação não implica necessariamente em uma relação de causalidade, ou seja, que um dos eventos tenha causado a ocorrência do outro. A presença de uma associação entre A e B não diz nada sobre a presença ou a direção da causalidade. Para demonstrar que A causou B (e não que B causou A nem que A e B foram ambos causados por C ou que seja a combinação das três situações), você necessita mais do que um coeficiente de correlação. A correlação pode ser apenas uma coincidência (acaso), ou seja, os dois eventos não têm qualquer relação.

Os autores expressaram os efeitos de uma intervenção/exposição em termos de provável benefício ou dano que se pode esperar? Quando o numerador corresponde a eventos que ocorrem de maneira prospectiva, como em estudos intervencionais e de coorte prospectiva, a medida de associação denomina-se risco relativo (RR). O RR corresponde à medida da força da associação entre a exposição e o evento (desfecho) e mede a incidência ou risco da doença (variável dependente) acontecer entre os expostos e os não expostos (à variável independente) em um período de tempo. Outras análises podem ser realizadas em estudos que avaliam o efeito de um tratamento, como: a Redução do Risco Absoluto (RRA), que representa a diferença do risco entre os expostos e não expostos e equivale a quantos eventos a menos se produzem ao expor ou não os sujeitos; a Redução do Risco Relativo (RRR), que é a diferença do risco entre os expostos e não expostos, mas com denominador constituído apenas pelos não expostos e; também é possível calcular o Número Necessário a Tratar (NNT), que equivalente ao número de indivíduos que se deve tratar para

evitar o evento indesejado. Quando o numerador corresponde a eventos que ocorreram, como em estudos de caso-controle e de coorte retrospectiva, a medida de associação denomina-se odds ratio (OR) e é definido como a probabilidade de que um evento ocorra dividido pela probabilidade de que ele não ocorra.

Os intervalos de confiança foram calculados e as conclusões dos autores os refletem? Por exemplo, o intervalo de confiança de 95% (IC95%) nos diz que há uma chance de 95% de que a diferença "real" dos resultados esteja entre os limites inferior e superior, ou seja, o resultado estará dentro daquele intervalo em 95 dos 100 estudos hipoteticamente realizados. Os 5 estudos excluídos deste intervalo representam valores extremos que possuem maior possibilidade de terem ocorrido por acaso. Quanto maior o estudo, mais estreito será o intervalo de confiança e, assim, maior a probabilidade de o resultado ser definitivo.

Além de um modelo estatístico adequado, a integridade e transparência dos resultados de um estudo dependem de como o valor de p selecionado será interpretado. A "significância estatística" muitas vezes é confundida com "importância científica". O valor de p se refere a um nível de significância que expressa a probabilidade máxima de se cometer um erro tipo I, ou seja, rejeitar uma H_0 que é verdadeira. O padrão científico geralmente adotado é de $p < 0,05$ como "estatisticamente significativo" e $p < 0,01$ como "estatisticamente muito significativo". Isso significa que o pesquisador assume o risco de que a rejeição da H_0 pode estar incorreta em 5% ou 1% dos casos, respectivamente. Um valor de p na faixa não significativa indica que não existe diferença entre os grupos ou que há poucos participantes para demonstrar essa diferença.

Tenha conhecimento dos dados referentes aos sujeitos da pesquisa. O cenário precisa estar bem descrito. Os grupos são realmente comparáveis? Para avaliar isso, é importante prestar atenção na tabela de caracterização da amostra, onde constam as características da linha de base dos grupos que estão sendo estudados, se são similares em termos de distribuição de idade e sexo, por exemplo. Se existirem diferenças importantes nestas características da linha de base, isso pode representar um desafio na interpretação dos resultados, pois gerará fatores de confusão.

Classificação dos desenhos de estudo

A escolha de um delineamento apropriado para um estudo é um passo crucial em qualquer pesquisa. Cada delineamento possui vantagens e desvantagens, além de potenciais fontes de viés e fatores de confusão. Cabe ao pesquisador, ou investigador, tentar reduzi-las ao elaborar o projeto de pesquisa. Os delineamentos de pesquisa podem ser classificados sob diversos aspectos:

Quanto à originalidade

Os estudos primários correspondem às investigações originais, como os ensaios clínicos randomizados, estudos transversais, estudos de coorte e estudos pré-clínicos. Os estudos secundários correspondem aos estudos formulados a partir da análise de informações de um conjunto de estudos primários, constituindo o nível mais alto de evidência. Incluem as revisões sistemáticas, estudos de análise econômica em saúde, relatórios de agências de avaliação de tecnologia em saúde e as diretrizes clínicas.

Quanto ao tipo de unidade de estudo

Os estudos clínicos envolvem, como modelo de pesquisa, pacientes, indivíduos, grupos ou populações, onde os investigadores designam pessoas elegíveis para grupos de intervenção ou avaliam alguma exposição. Os principais estudos estão representados na figura 6.

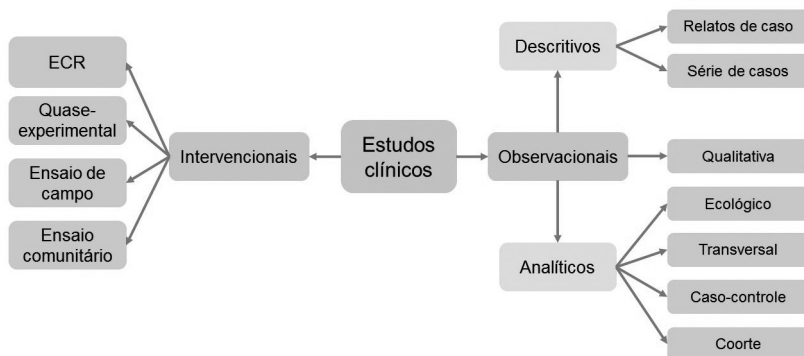


Figura 6. Representação esquemática dos estudos clínicos.

Os estudos experimentais (Figura 7) envolvem modelos com animais de experimentação (ratos, camundongos, coelhos, porcos, etc.), modelos biológicos (cultura de células e tecidos) e modelos físicos e/ou mecânicos (pulmão, coração, softwares, simuladores, etc.).

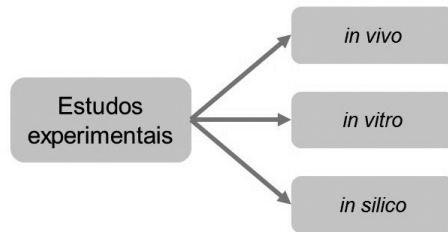


Figura 7: Representação esquemática dos estudos experimentais.

Quanto à inferência

Os estudos clínicos podem ser classificados em observacionais, em que os casos são observados pelo pesquisador quanto às características da doença ou transtorno e sua evolução, sem intervir ou modificar qualquer aspecto que esteja estudando, ou seja, o investigador mede, mas não intervém e; intervencionais, também chamados de experimentais, onde o pesquisador não somente observa, mas também interfere pela exclusão, inclusão ou modificação de um determinado fator.

Nos estudos intervencionais, o pesquisador tem a possibilidade de controlar as condições experimentais e a constituição dos grupos de indivíduos a serem comparados. Quando há necessidade de comparação, a alocação dos sujeitos deve ocorrer de forma aleatória, por meio de uma randomização. Os principais estudos experimentais são ensaios clínicos randomizados (cujos participantes são os pacientes), ensaios de campo (em que os participantes são pessoas saudáveis) e ensaios comunitários (onde os participantes são os próprios membros da comunidade).

Quanto ao perfil de avaliação

Os estudos observacionais podem ser divididos em descritivos e analíticos. Frequentemente, o primeiro passo de uma pesquisa é a simples descrição da situação. No entanto, deve-se ter cuidado, pois o fato de descrever a amostra do estudo não torna o estudo puramente descritivo. Os estudos descritivos se limitam a descrever a caracterização de aspectos semiológicos, etiológicos, fisiopatológicos e epidemiológicos de uma doença, nova ou rara, em uma população. São utilizados para conhecer a sua distribuição no tempo, no espaço e conforme características individuais. Podem abranger desde relatos ou séries de casos até estudos populacionais, como estudos ecológicos. Importante: os estudos descritivos não testam hipóteses, sendo o primeiro passo de uma investigação.

Os estudos analíticos são modelos de estudos utilizados para verificar uma hipótese. O pesquisador introduz um fator de exposição ou um novo recurso terapêutico, e avalia-o utilizando ferramentas estatísticas. A precisão dos resultados está diretamente associada a representatividade dos dados na população, o que determina a importância do cálculo adequado da amostra. Estimam incidência e prevalência e, geralmente, constituem-se de estudos primários.

Quanto ao período de seguimento

Um estudo longitudinal, também denominado seguimento, sequencial ou *follow up*, é um estudo onde existe uma sequência temporal (ao longo do tempo) conhecida entre uma exposição, ausência da mesma ou intervenção, e o aparecimento da doença ou fator. Portanto, é um filme da situação. Em geral, são mais caros e há expressiva perda de seguimento. Os estudos longitudinais podem ser prospectivos ou retrospectivos.

Os estudos transversais (seccionais ou de prevalência), por sua vez, são estudos em que as medidas de exposição e efeito (doença) são realizadas ao mesmo tempo. Portanto, apresenta-se como uma fotografia ou corte instantâneo que se faz numa população por meio de uma amostragem, examinando-se a presença ou ausência da exposição e a presença ou ausência do efeito ou da doença. Possui como principais vantagens o fato de

serem de baixo custo, fáceis de conduzir e por praticamente não apresentarem perdas.

Para diferenciar os estudos em longitudinais ou transversais, é importante considerar o impacto do fator tempo sobre o estudo. É preciso avaliar se a mudança temporal foi utilizada somente como instrumento de estudo (transversal), ou se foi como objeto do estudo (longitudinal).

Quanto à direção do tempo

Podemos classificar os estudos conforme a direcionalidade temporal, isto é, o período de tempo durante o qual os dados foram registrados em relação ao tempo no qual o estudo começou, em prospectivo (contemporâneo, prospectivo concorrente ou concorrente) ou retrospectivo (histórico, prospectivo não concorrente, não concorrente ou invertido).

O estudo é prospectivo quando se elabora o estudo no presente e o mesmo é seguido para o futuro. O termo "prospectivo" refere-se ao momento da coleta dos dados e não a relação entre exposição e efeito (doença). Neste caso, parte-se da causa para o efeito. O estudo é retrospectivo quando se realiza o estudo a partir de registros existentes do passado, como por exemplo, por meio de prontuários eletrônicos. Para isso, é fundamental que haja credibilidade nos dados de registros a serem computados, em relação à exposição do fator e/ou à sua intensidade, assim como pela ocorrência da doença ou situação clínica. Assim, parte-se do efeito para a causa.

Quanto à existência de controle ou comparador

Pode-se ter estudos não controlados, que se limitam aos relatos de caso, séries de casos, estudos de "antes e depois" e de relação estímulo/efeito. Trata-se dos estudos onde se registra os dados relativos à observação de grupos de indivíduos portadores de uma doença, sem utilizar um grupo controle ou placebo. Comumente são utilizados em estudos descritivos (relatos e séries de caso) para verificar a frequência de doenças e, assim, permitir ao pesquisador formular hipóteses a serem confirmadas por meio de estudos analíticos.

Os estudos controlados são estudos que envolvem um grupo de casos e um grupo controle. O grupo controle deve ter seus sujeitos o mais semelhante possível daqueles do grupo de casos. Diferem apenas pelo fato de que o primeiro vai receber o placebo ou o tratamento padrão ou ausência de uma doença ou fator de exposição, enquanto que o segundo receberá um novo tratamento ou fator de exposição ou presença da doença.

Os estudos comparativos comparam grupos diferentes, não sendo necessariamente um o controle do outro. Por exemplo, um estudo que visa comparar duas técnicas cirúrgicas para tratamento da obesidade, tendo como desfecho a perda de peso corporal. Ao comparar os tipos de cirurgia, não se trataria de um estudo controlado, e sim de um estudo comparativo.

Além disso, existe um tipo especial de controle para seguimento de uma condição ou de uma intervenção, que exclui os fatores das variáveis interpessoais, o chamado estudo autocontrolado (autopareado ou *self paring*). Por exemplo, um estudo que busca verificar as repercussões biológicas locais e/ou sistêmicas de um implante de silicone com diferentes características físicas e/ou químicas de cada lado do corpo de um modelo animal experimental. Esse modelo de estudo também pode ser autocomparativo.

Quanto aos tipos de frequência

Os estudos de prevalência (detecção de casos, rastreamento ou *screening*) referem-se à proporção de casos (doença ou condição) existentes na população em um determinado momento, constituindo-se em estudos transversais. Os estudos de incidência referem-se à proporção de casos novos (doença) ou de desfechos novos dos casos existentes de uma doença específica, em um grupo inicialmente livre da condição clínica e que a desenvolve depois de um período determinado de tempo. Utiliza-se a medida de incidência como uma estimativa de risco. Constituem-se em estudos longitudinais.

Os estudos de acurácia são utilizados para avaliar o desempenho de um novo teste diagnóstico (teste índice), garantindo que seu resultado positivo indique, realmente, a

presença de uma doença, e que seja negativo na ausência da mesma. Para tal, utiliza-se um padrão de referência (“padrão-ouro” ou *gold standard*). A validação (acurácia) do novo teste diagnóstico é dada através de parâmetros de sensibilidade (ser positivo em doentes) e de especificidade (ser negativo em não doente). Trata-se de um estudo do tipo transversal.

Os estudos de validade, fidedignidade e concordância estão relacionados às propriedades de avaliação de um instrumento de pesquisa. A validade (ou acurácia) investiga até que ponto o método realmente mede o que se propõe a medir. Pode ser medida através de uma curva ROC (*Receiver Operator Characteristic*) (Figura 8). A concordância serve para avaliar o quanto dois métodos, que supostamente se propõem a aferir a mesma medida, são parecidos (concordantes). Pode ser avaliado através de um coeficiente do índice Kappa (K). Quando há concordância total entre os observadores, o coeficiente K é igual a 1 e, se há discordância total, o índice K é igual a zero. A fidedignidade reflete o erro, tanto aleatório como sistemático, inerente a qualquer medida. Investiga o grau de conformidade entre múltiplas medidas (reprodutibilidade) de uma mesma variável em um mesmo objeto ou indivíduo. Um método fidedigno não deve produzir resultados significativamente diferentes se for repetido com os mesmos sujeitos, ou em tempos diferentes ou por diferentes observadores.

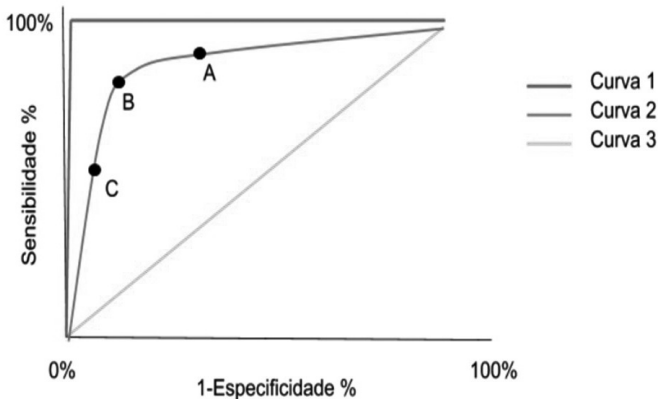


Figura 8. Curva ROC.

Quanto à procedência da equipe

Pode-se ter um estudo de centro único, onde os integrantes da pesquisa pertencem a apenas uma instituição de pesquisa, universitária ou não, ou multicêntrico, em que temos um estudo de caráter cooperativo entre diversas instituições. Um aspecto importante dos estudos multicêntricos é que ocorram de forma padronizada.

Estudo piloto

Não é propriamente um delineamento de estudo, e sim uma ferramenta que fornece informações para um melhor planejamento de um estudo maior. Através de um estudo piloto, é possível avaliar a viabilidade do projeto, detectar possíveis falhas no método de pesquisa e na elaboração e aplicação de um instrumento. O principal objetivo do estudo piloto é fornecer uma base para o cálculo amostral de um estudo maior, quando não há evidências disponíveis para o cálculo.

Hierarquia dos níveis de evidência

A hierarquia dos níveis de evidência é baseada no grau de confiança da informação, com base no delineamento do estudo. A pirâmide tradicional das evidências é amplamente conhecida (Figura 9), onde as revisões sistemáticas encontram-se posicionadas no topo, seguidas pelos ensaios clínicos randomizados. A seguir, os estudos observacionais como estudos de coorte, de caso-controle, transversal e ecológico, seguidos de série de caso, “opiniões de especialistas” e, na parte inferior (base), a pesquisa básica. A pesquisa qualitativa não se enquadra nesta hierarquia particular de evidências.



Figura 9. Pirâmide tradicional de evidência científica.

No entanto, é importante ressaltar que nem todas as evidências são, necessariamente, equivalentes. Ou seja, mesmo que revisões sistemáticas e metanálises sejam consideradas como os estudos de maior nível de evidência, erros de delineamento e limitações podem comprometer a confiabilidade das informações. Falhas metodológicas graves e fundamentais comumente ocorrem mesmo em artigos publicados em periódicos de “qualidade”, com alto fator de impacto.

Recentemente, uma revisão da pirâmide tradicional foi proposta por Murad *et al.*, 2016, na qual as linhas que separam os desenhos do estudo passam a ser onduladas (qualidade da evidência) e as revisões sistemáticas são removidas da pirâmide (Figura 10A e B). A nova pirâmide propõe que as revisões sistemáticas funcionem como uma lente através da qual os estudos (evidências) são vistos (ou seja, avaliados e aplicados) (Figura 10C)

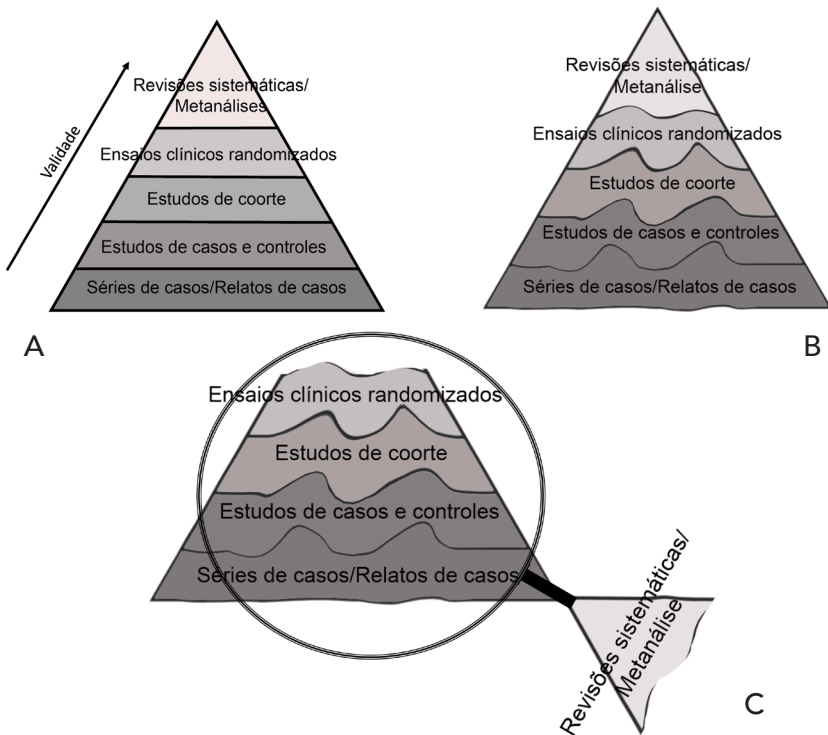


Figura 10. Nova pirâmide de evidência (adaptado de Murad *et al.*, 2016).

Considerações éticas

Todos os estudos envolvendo seres humanos devem ser revisados e aprovados por um comitê de ética em pesquisa (CEP), através da submissão na Plataforma Brasil, seguindo os princípios do consentimento informado, da confidencialidade dos dados coletados, do respeito pelos direitos humanos e da integridade científica. Quando os aspectos éticos das pesquisas se enquadram nas áreas temáticas especiais, como Genética Humana e Biossegurança, compete à Comissão Nacional de Ética em Pesquisa (CONEP) a apreciação ética. Também em caso de dúvida, o CEP pode e deve enviar à CONEP o caso para posicionamento.

Segundo a Resolução 466/2012, considera-se "(...) pesquisa em seres humanos as realizadas em qualquer área do conhecimento e que, de modo direto ou indireto, envolvam indivíduos ou coletividades, em sua totalidade ou partes, incluindo o manejo de seus dados, informações ou materiais biológicos." Envolve também "(...) as entrevistas, aplicações de questionários, utilização de banco de dados e revisões de prontuários." Dados de domínio público dispensam apreciação pelo CEP, mas pode-se solicitar um Parecer de Dispensa. Também não passam por avaliação os estudos de gestão e melhoria de processos, de validação de técnicas e de revisão sistemática.

Os participantes da pesquisa devem assinar o Termo de Consentimento Livre e Esclarecido (TCLE), concordando com a participação do estudo. Os princípios da autonomia, beneficência, não maleficência, justiça e equidade devem ser seguidos na elaboração do TCLE e na condução do estudo. Aos participantes da pesquisa fica preservado o direito de abandonar o estudo em qualquer momento, sem qualquer prejuízo. Os pesquisadores também devem atentar quais os riscos e benefícios do estudo, procurando identificar possibilidades de reduzir possíveis riscos aos participantes.

A prática que envolve o uso de animais é regulada pelo Conselho Nacional de Controle de Experimentação Animal (CONCEA), que orienta as Comissões de Ética no Uso de Animais (CEUA), pesquisadores e docentes na utilização de animais em aulas e pesquisas. A Lei nº 11.794/2008 (Lei Arouca) "(...) estabelece que

haja procedimentos bem definidos para produção, manutenção ou utilização de animais para atividades de ensino e pesquisa em todo o território nacional." Além disso, cabe monitorar e avaliar a introdução de métodos alternativos (como modelos *in vitro* e *in silico*), substituindo ou reduzindo o número de animais usados em experimentos. Cabe a apreciação ética para as atividades de ensino e pesquisa que envolvem animais vertebrados ou materiais e dados derivados destes ou parte destes, não sendo aplicáveis quando forem utilizados, unicamente, os animais invertebrados. Entretanto, no caso de os invertebrados serem obtidos a partir de animais, deve ser encaminhada proposta à CEUA mesmo que o objeto de estudo sejam os invertebrados, pois há envolvimento de animais vertebrados na avaliação.

As pesquisas que envolvem a utilização de animais devem ser adequadamente justificadas. Para minimizar o seu uso, vale destacar o princípio dos 3 R's: "*replace*" (substituição do uso de animais por métodos alternativos), "*reduce*" (redução do número de pesquisas realizadas em modelos animais, redução do número de animais utilizados e aumento na qualidade do tratamento estatístico) e "*refine*" (refinamento das técnicas utilizadas, visando diminuir a dor e o sofrimento dos animais, incluindo cuidados de analgesia e assepsia nos períodos pré-, trans- e pós-operatórios). No entanto, a utilização desses princípios deve ter o cuidado para não comprometer a validade do estudo.

As questões éticas são muito importantes e dizem respeito a todas as ações desenvolvidas na condução da pesquisa. Quando você estiver lendo um artigo científico, assegure-se de que o estudo que você está lendo recebeu aprovação ética. Lembre-se que os periódicos rotineiramente se recusam a publicar pesquisas que não tenham sido avaliadas e aprovadas por um comitê de ética.

Referências

Bonita R, Beaglehole R, Kjellström T. Epidemiologia básica. 2. ed. São Paulo: Santos, 2010. 213 p.

Brasil. Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia.

Por que pesquisa em saúde? / Ministério da Saúde, Secretaria de Ciência, Tecnologia e Insumos Estratégicos, Departamento de Ciência e Tecnologia. – Brasília : Ministério da Saúde, 2007. 20 p.

Constitution of the World Health Organization. New York. World Health Organization, 1946.

Callegari-Jacques SM. Bioestatística: princípios e aplicações. Porto Alegre: ArtMed, 2011. 255 p.

Greenhalgh T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5a ed. Porto Alegre: Artmed, 2015. 376 p.

Hochman B, Nahas FX, de Oliveira Filho RS, Ferreira LM. Desenhos de pesquisa. Acta Cir. Bras. 2005;20(2):1-8.

Lei N° 11.794, de 8 de outubro de 2008. Disponível em http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/l11794.htm

Mancuso ACB, Nune LN, Guimarães LSP, Camey AS, Hirakata VN. Métodos Estatísticos para Desfechos Quantitativos utilizando o SPSS v.18, 2017. Disponível em <https://sites.google.com/a/hcpa.edu.br/bioestatcursos/home/ii>. Acessado em 13/07/2020.

Murad NA, Alsawas AM, Alahdab F. New evidence pyramid. 2016;21(4):125-7.

Oliveira Filho PF. Epidemiologia e bioestatística: fundamentos para a leitura crítica. Rio de Janeiro (RJ): Rubio, 2015. 221 p.

Resolução N° 466, de 12 de dezembro de 2012. Disponível em <http://www.conselho.saude.gov.br/resolucoes/2012/Reso466.pdf>

Análise crítica e qualidade metodológica de artigos científicos

*Hellen Meiry Grosskopf Werka
Pedro Henrique Comerlato
Edison Capp
Otto Henrique Nienov*

O que é pesquisa científica?

A pesquisa científica pode ser definida como uma investigação focada e sistemática com o objetivo de gerar novos conhecimentos, que faz parte do método científico. O método científico, por sua vez, é o processo rigoroso que utiliza a abordagem sistemática, empírica, controlada e crítica, ou seja, existe uma ordem, uma disciplina, um rigor que confere confiabilidade e uma sequência lógica para resolver a questão de pesquisa e gerar um novo conhecimento.

Partindo de uma observação, de uma maneira geral, formula-se uma pergunta de pesquisa, a partir da qual ocorre a revisão da literatura e a formulação de hipóteses. Em seguida, testa-se a hipótese por meio de um experimento, faz-se a análise dos dados e divulga-se os resultados, estejam alinhados ou não com a hipótese. Logo, é necessário que o objetivo da pesquisa seja claro, tangível e passível de ser respondido.

Em um projeto de pesquisa temos objetivos primário e secundários. O objetivo primário ou principal reflete a pergunta de pesquisa. Os objetivos secundários ou adicionais estão relacionados aos efeitos adicionais do estudo. A partir dos objetivos, são extraídas as variáveis, que representam as características de interesse de elementos individuais de uma população ou amostra pesquisada.

As variáveis são parte estruturante de uma hipótese relacionada ao objetivo. As variáveis podem ser classificadas,

conforme a natureza estatística dos dados, em qualitativas ou quantitativas ou, conforme a natureza da informação, em primárias e secundárias. A variável primária, também chamada de variável alvo ou desfecho, é aquela capaz de proporcionar a evidência clínica mais relevante e convincente em relação ao objetivo primário. As variáveis secundárias são aquelas adicionais, relacionadas aos objetivos secundários.

O teste de hipótese verifica se são verdadeiras as afirmações sobre os parâmetros de uma ou mais populações. O delineamento de um estudo deve ser iniciado pela formulação da hipótese nula (H_0), ou seja, a hipótese que se quer rejeitar, com um nível de significância (α). Já a hipótese que o estudo busca comprovar corresponde à hipótese alternativa (H_A). Por serem baseados em estatísticas de uma amostra realizada em uma população, os testes estão sujeitos a erros que podem ser do tipo I (rejeita-se H_0 quando esta é verdadeira, sendo denotada por α) ou do tipo II (não se rejeita H_0 quando esta é falsa, sendo denotada por β) (Figura 1).

		REALIDADE	
		VERDADEIRO	FALSO
RESULTADO	POSITIVO	OK	ERRO TIPO II
	NEGATIVO	ERRO TIPO I	OK

Figura 1. Teste de hipóteses.

Por exemplo, um estudo quer avaliar se o nível de atividade física difere entre indivíduos com e sem hipertensão. Logo:

H_0 : o nível de atividade física não difere entre os sujeitos hipertensos e não hipertensos.

H_A : o nível de atividade física difere entre os sujeitos hipertensos e não hipertensos.

Se na realidade o nível de atividade física não difere entre os grupos (hipertenso e não hipertenso), mas no estudo rejeita-se a H_0 pois é encontrada uma diferença, têm-se um erro do tipo I, ou seja, o estudo gerou um resultado falso positivo. Do contrário, se na realidade o nível de atividade física difere entre os grupos, mas no estudo não se rejeita a H_0 pois não é encontrada uma diferença entre os grupos, têm-se um erro do tipo II, ou seja, o estudo obteve um resultado falso negativo.

Medicina Baseada em Evidências

A Medicina Baseada em Evidências engloba a melhor evidência científica disponível, o conhecimento clínico e as preferências do paciente para a tomada de decisão. Ou seja, é muito mais do que apenas ler artigos científicos. A Medicina Baseada em Evidências envolve:

- 1) formulação do problema (converte-se a necessidade de informações em questões de pesquisa);
- 2) busca de respostas em bases de dados (realiza-se o rastreio das evidências disponíveis);
- 3) análise crítica das evidências encontradas (verifica-se o quão próximas as evidências estão da verdade);
- 4) implementação dos resultados (seja na prática clínica, no ambiente ou nos pacientes) e;
- 5) avaliação do impacto (verifica-se o impacto gerado pela implementação dos resultados).

Resumindo, a abordagem baseada em evidências para tomada de decisão resulta em investigação de respostas de modo sistemático com fins de implicações práticas. Desta forma, deve-se ler os artigos certos da forma certa e, a partir disso, modificar o comportamento.

Por que ler um artigo científico?

Alguns dos gatilhos que levam a leitura de artigos científicos são: a curiosidade, o surgimento de algum problema ou

a formulação de uma pergunta de pesquisa. Antes de se iniciar a busca de um artigo, é essencial formular uma pergunta científica. A pergunta deve ser interessante e motivar o pesquisador e sua equipe. Na pergunta, é importante observar os seguintes fatores: qual seria a população a que se aplica a pergunta? Há algum fator de exposição ou alguma intervenção terapêutica? Existe um grupo comparador ou uma não exposição (fator)? Qual o resultado de interesse? Esses fatores podem ser demonstrados através do acrônimo PICO (Quadro 1).

Quadro 1. Descrição do acrônimo PICO.

Acrônimo	Significado
P	Refere-se ao paciente, problema ou população. Com quem a pergunta se relaciona?
I	Diz respeito à intervenção, indicação, interesse ou exposição. Pode ser uma terapia, um teste diagnóstico, questão ou fator.
C	Indica um procedimento padrão, comparador, placebo, não-intervenção ou não-exposição. Pode ser outra intervenção, teste diagnóstico, placebo, padrão-ouro ou ausência do fator.
O	Corresponde ao desfecho ("outcome"). É o resultado de interesse ou esperado.

Nas pesquisas qualitativas, há também o acrônimo PICo. Porém, neste, as letras recebem outro significado (Quadro 2).

Quadro 2. Descrição do acrônimo PICo para pesquisa qualitativa.

Acrônimo	Significado
P	Refere-se ao paciente, problema ou população. Com quem a pergunta se relaciona?
I	Diz respeito ao interesse. Qual o objetivo da pesquisa? O que será observado?
Co	Corresponde ao contexto. Qual o ambiente ou situação específica?

Uma pergunta de pesquisa adequada possibilita a definição correta de que informações (evidências) são necessárias para respondê-la. Os termos utilizados na pergunta de pesquisa podem ser chamados de descritores e farão parte da estratégia de busca. Os descritores são termos organizados em estruturas hierárquicas e facilitam a pesquisa e posterior recuperação de um artigo. As palavras-chaves são termos retirados de textos de linguagem livre, aleatórios, e que não obedecem a nenhuma estrutura.

É importante observar os descritores cadastrados nas bases de dados e as palavras-chaves utilizadas nos artigos científicos. Por exemplo, na Pubmed, os descritores podem ser buscados na lista de termos "Mesh". No caso da Lilacs, utiliza-se os descritores em ciências da saúde (Decs). Quanto ao uso de abreviaturas, é aconselhável verificar no periódico qual a melhor forma de uso de abreviaturas. Em alguns casos, há siglas comumente utilizadas que não precisam ser descritas, apenas abreviadas. Por exemplo, o *International Journal of Obesity* descreve o modo de apresentação das abreviaturas mais comuns que não necessitam estar definidas. Além disso, o periódico pode fornecer recomendações para análise estatística, apresentação de unidades de medida e utilização de terminologias.

Para busca nas bases de dados, deve-se primeiro realizar a busca pelos descritores isoladamente. Em seguida, deve-se combinar ou delimitar esses descritores com o uso de operadores booleanos: "AND" (para combinação restritiva), "OR" (para combinação aditiva) e "NOT" (para condição excludente) (Figura 2).

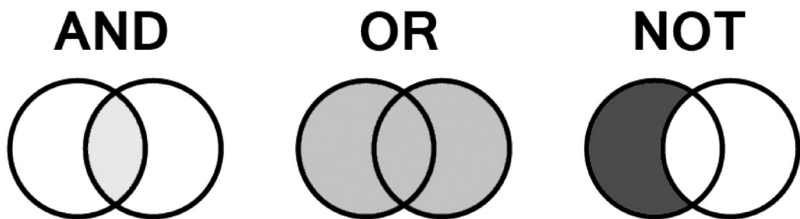


Figura 2. Operadores booleanos e delimitações.

A busca deve compreender, preferivelmente, mais de uma base de dados. Uma busca simples no *PubMed* não será suficiente. Mas, é um bom começo. É recomendado iniciar a busca pela base de dados mais importante para o tema da pesquisa e, se possível, que seja a mais familiar para o pesquisador. Vale destacar que cada base de dados possui estratégias específicas de busca, como truncagens e explosões de termos. Saber como acessar e buscar nas bases de dados, traz economia de tempo e uma melhor capacidade de encontrar toda a evidência disponível.

Para executar uma boa busca, o pesquisador precisa vencer algumas barreiras, como falta de tempo, de habilidades de pesquisa, de domínio nas bases de dados, além do excesso de informações, seja pelo volume ou pela complexidade do tema. Existem profissionais capacitados que podem auxiliar a enfrentar essas dificuldades, como o bibliotecário e especialistas da área. Tutoriais online de busca nas bases de dados também podem auxiliar.

Há ainda a literatura “cinzenta”, que corresponde a estudos realizados, porém não indexados. Na revisão sistemática essa literatura deve ser incluída na estratégia de busca, pois é uma excelente fonte de estudos com resultados negativos ou não significativos, reduzindo assim o viés de publicação. Idealmente, toda a literatura sobre a questão de pesquisa deve ser revisada, mas pode-se parar de realizar a busca quando encontramos a resposta da nossa pergunta de pesquisa. Vale destacar que, durante essa busca, encontraremos as lacunas do conhecimento onde surgirão novas perguntas de pesquisa.

Estrutura de um artigo científico

Para saber ler um artigo científico é importante conhecermos a sua estrutura. O formato padrão constitui-se de: título, resumo, introdução, método, resultados, discussão e referências.

Um bom título deve informar de forma clara a que se destina o artigo. Após o título, normalmente, encontram-se ordenados os autores e suas respectivas afiliações e contatos. Um autor é um indivíduo que fez uma contribuição intelectual significativa para o estudo. Podemos destacar quatro critérios básicos para identificar um autor:

- 1) ter contribuído substancialmente com a concepção e desenho do estudo, aquisição de dados, análise e interpretação dos resultados;
- 2) ter elaborado ou revisado o artigo;
- 3) ter aprovado a versão final e;
- 4) ter concordado em ser responsável por todos os aspectos relacionados à precisão ou integridade de qualquer parte do trabalho.

A ordem de autoria deve ser uma decisão conjunta dos coautores. Indivíduos envolvidos em um estudo, mas que não satisfazem os critérios de autoria, devem ser listados como “colaboradores” ou “indivíduos reconhecidos” ou mencionados nos “agradecimentos”. Para ensaios grandes e multicêntricos, a lista de pesquisadores e centros participantes normalmente é publicada juntamente com uma declaração das contribuições individuais feitas.

O resumo é um texto curto, com cerca de 250 palavras, em que, com a simples leitura, o leitor consegue identificar os pontos-chaves do estudo. Portanto, o resumo é um texto, com uma quantidade predeterminada de palavras, onde se expõe o objetivo do artigo, a metodologia utilizada para solucionar o problema e os resultados alcançados. Seguido, temos as palavras-chaves, que são palavras características do tema de estudo.

A introdução tem o objetivo de situar o leitor, de modo resumido, no contexto do tema pesquisado. Além de esclarecer quais são as bases da pesquisa, a introdução deve afirmar claramente a pergunta de pesquisa (objetivo) e/ou a hipótese que os autores decidiram testar, além das justificativas que motivaram o pesquisador a realizar o estudo. Geralmente a pergunta de pesquisa é apresentada no final da introdução e retomada no primeiro parágrafo da discussão, seguido da resposta que o estudo encontrou.

O método informa sobre a maneira como foi realizado o estudo e como os resultados foram analisados. Normalmente, essa seção é dividida em itens apresentados de forma cronológica do

desenvolvimento da pesquisa. São informados o delineamento do estudo, as características da amostra, os critérios de inclusão e de exclusão, os instrumentos e experimentos realizados, a análise estatística e o cálculo do tamanho da amostra, além dos aspectos éticos.

Os resultados reportam os achados, sumarizados por tabelas, gráficos e figuras. Na discussão, é feita a avaliação e argumentação destes resultados, ou seja, o que significam. Nesta parte do artigo, o autor deve fazer uma exposição e uma discussão das teorias que foram utilizadas para entender e esclarecer o problema, apresentando-as e relacionando-as com a pergunta de pesquisa. Também é importante destacar as limitações e forças do estudo. Após, são apresentadas as conclusões, evidenciando com clareza e objetividade a resposta ao problema do estudo.

Por fim, as referências correspondem ao conjunto de documentos, impressos ou eletrônicos, utilizados como referência para a elaboração do artigo científico. As publicações devem ter sido mencionadas no texto.

Como escrever um artigo?

Escrever é o que torna a ciência real e utilizável para os outros. A redação do artigo científico é fundamental para a divulgação dos resultados e, assim, alcançar um impacto na comunidade científica. Para escrever um artigo científico, deve-se utilizar uma linguagem objetiva, precisa e com vocabulário técnico. O estilo de escrita deve ser coerente, conciso e fiel às fontes citadas. São importantes também a correção gramatical e o uso dos recursos ilustrativos.

A escrita ocorre em várias etapas e inclui muitas revisões. A revisão do que foi escrito é um ótimo exercício, pois possibilita identificar potenciais erros ou formas de melhorar a apresentação dos dados. Infelizmente, nem todas as boas ideias são executadas da melhor maneira. As principais falhas que levam a rejeição de artigos incluem:

- 1) o estudo não ter abordado um aspecto científico importante;
- 2) o estudo não foi original;

- 3) o estudo realmente não testava a hipótese do autor;
- 4) um delineamento diferente deveria ter sido utilizado;
- 5) dificuldades metodológicas comprometeram o protocolo original do estudo;
- 6) o tamanho da amostra foi insuficiente para o objetivo do estudo;
- 7) o estudo não foi ou foi inadequadamente controlado;
- 8) os grupos não eram comparáveis;
- 9) a análise estatística estava incorreta ou inapropriada;
- 10) os autores tiraram conclusões injustificadas a partir de seus dados ou a conclusão não respondia a pergunta da pesquisa;
- 11) existia um conflito de interesses ou;
- 12) na pior das possibilidades, o artigo foi tão mal escrito que se tornou incompreensível.

Dúvida: ler ou não um artigo

Se sua dúvida é ler ou não um artigo científico, comece observando o método. É possível descartar um artigo antes mesmo de olhar os resultados e deixar-se influenciar pelos achados. Lembre-se: há boas razões pelas quais os autores do estudo não realizaram um estudo da melhor forma: pergunta de pesquisa mal formulada, falta de planejamento, problemas de condução, instrumento inadequado, vieses, fatores de confusão, tratamento estatístico inapropriado, amostra insuficiente. Da mesma forma, eles sabem tão bem quanto você que seu trabalho seria mais válido cientificamente se esta ou aquela dificuldade, prevista ou não, não tivesse surgido no decorrer do estudo.

A validade científica, originalidade e importância dos estudos são avaliados antes de serem publicados, através de uma revisão por pares (peer review). Durante a revisão por pares, os revisores observam todos os fatores potencialmente problemáticos. A revisão é um processo positivo, que serve para apontar falhas ou pontos de melhoria, garantindo que periódicos publiquem ciência de boa qualidade e que traga benefício para toda a comunidade científica. É um trabalho gratuito executado por pesquisadores voluntários, em geral com vasta experiência no

campo da pesquisa. Através da revisão, busca-se atingir trabalhos mais robustos, mais fáceis de ler e mais úteis.

Da mesma forma, precisamos desenvolver essa análise crítica com os nossos trabalhos e com a avaliação da qualidade metodológico dos trabalhos que iremos encontrar ao longo da nossa trajetória. Ao ler um artigo científico, questione: qual foi a questão de pesquisa e porque ela foi necessária? Qual foi a hipótese testada? Qual foi o delineamento do estudo? O tipo de estudo foi apropriado para o objetivo da pesquisa? Ou seja, é capaz de responde-lo? Desta forma, busque identificar o acrônimo PICO. Verifique se é um estudo primário (original) ou secundário. Identifique os objetivos, as hipóteses, o delineamento do estudo e analise se o delineamento é adequado para os objetivos propostos (Quadro 3).

Quadro 3. Delineamentos indicados de acordo com o objetivo de estudo (adaptado de Greenhalgh, 2015).

OBJETIVO	DESCRIÇÃO	DELINEAMENTO
Causação (etiologia e etiopatogenia)	Determinação se uma exposição (agente) está relacionada ao desenvolvimento de doença.	Estudos de coorte ou caso-controle ou relatos de casos
Diagnóstico	Demonstração se um novo teste diagnóstico é válido (validade) e confiável (reprodutibilidade).	Estudos transversais
Estudos psicométricos	Mensuração de atitudes, crenças ou preferências, frequentemente a respeito da natureza da doença ou de seu tratamento.	Pesquisa qualitativa
Prognóstico	Determinação do que provavelmente aconteceria a alguém cuja doença fosse detectada em estágio inicial.	Estudos longitudinais
Rastreamento	Demonstração do valor de testes que podem ser aplicados a grandes populações, bem como detectar a doença em um estágio pré-sintomático.	Estudos transversais
Tratamento	Testagem da eficácia de tratamentos medicamentosos, procedimentos cirúrgicos, métodos alternativos de prestação de serviços ou outras intervenções.	Ensaio Clínico Randomizado

Avaliando a qualidade metodológica

A seguir, são apresentadas seis questões essenciais que devem formar a base de sua decisão sobre dispensar, confiar ou olhar com cautela para um artigo científico.

O estudo apresenta originalidade?

Não há motivos para testar uma hipótese de um tema que já está amplamente comprovado ou desgastado. Na vida real, nem todo estudo é inédito. Mas, é possível adicionar esse estudo como uma parte num contexto científico maior, como se fosse uma peça num quebra-cabeças. Nos casos em que há estudos similares, pergunte-se se essa nova pesquisa acrescenta algo à literatura, de alguma forma. Por isso, antes de executar uma ideia, procure investigar nas evidências disponíveis (e isso inclui a literatura cinzenta), se o tema de pesquisa está suficientemente estabelecido e se sua proposta apresenta originalidade.

Neste aspecto, inclusive, é possível encontrar as justificativas para a realização do estudo, como: o estudo é maior, tem maior duração ou é mais substancial do que o(s) anterior(es); a metodologia é mais rigorosa ou utiliza um instrumento alternativo; os resultados numéricos desse estudo representam um acréscimo significativo a uma metanálise de estudos anteriores; a população estudada é de algum modo diferente (por exemplo, diferentes grupos étnicos, idades, massa corporal ou sexos em relação a estudos anteriores) ou; existe uma necessidade além do aspecto científico que motiva a pesquisa, como uma circunstância social ou política.

O objetivo do estudo é claro e apropriado?

Se a pergunta de pesquisa não está contemplada no objetivo, não podemos chegar a nenhuma conclusão. Observe se o objetivo traduz facilmente o título do trabalho. Caso contrário, o título pode induzir o leitor à conclusão que sequer está sendo analisada pelo estudo. Da mesma forma, verifique se o teste de hipótese gerado no estudo, foi elaborado a partir do título e do objetivo. Nesse momento, é possível encontrar muitos erros de interpretação e de conclusão equivocadas. A partir do objetivo, indague se o tipo de estudo foi o melhor delineamento para atender o estudo.

Antes de examinar os resultados, tente descrever quais seriam as melhores variáveis de desfecho. Talvez o autor poderia responder ao objetivo de outra forma ou outra variável poderia mostrar melhor o desfecho de interesse. Considere o delineamento do estudo e pense nos potenciais vieses e limitações metodológicas. Também verifique se os fatores de confusão foram considerados, por exemplo, se foram incluídos no controle da análise, por meio da regressão multivariada, ou então consideradas nos critérios de elegibilidade do estudo.

O estudo é sobre quem?

Nem sempre os resultados obtidos numa amostra são condizentes com uma população maior (validade externa). Alguns pontos a serem observados quanto a essas diferenças são: estado de doença (mais ou menos grave); características étnicas e regionais; hábitos diários (estilo de vida); cuidados de atenção (pesquisa versus vida real); distintas linhagens de célula e modelos animais; ambiente controlado.

Deve-se observar o modo como os participantes da pesquisa foram recrutados para o estudo, se não houve viés de seleção. Por exemplo, ao divulgar em um jornal a busca de pacientes para um estudo, pode-se atrair apenas pessoas motivadas e que leem jornais. A divulgação deve incluir meios amplos e alternativos de mídia, para alcançar o máximo de representatividade da população.

Também é necessário analisar os critérios de inclusão e exclusão definidos pelo estudo. Os critérios de inclusão refletem as características-chave da população-alvo que os pesquisadores utilizarão para responder à pergunta do estudo. Os critérios de exclusão correspondem a características adicionais dos sujeitos incluídos que poderiam interferir no desfecho do estudo. Lembre-se: um participante da pesquisa somente pode ser excluído após ter sido incluído no estudo.

Os critérios de inclusão e exclusão fazem parte da prática padrão de um estudo e necessária na elaboração de projetos de pesquisa. Caracterizar a amostra adequadamente e em circunstância de vida real viabiliza maior validade externa. Por

isso, outro ponto necessário a se ter cuidado é se esse critério pré-definido é apropriado especialmente para a validade externa.

A seguir, são apresentados alguns exemplos de erros comuns de critérios de inclusão e exclusão:

Exemplo 1: Utilizar a mesma variável para definir critérios de inclusão e de exclusão. Por exemplo, critério de inclusão ser do sexo feminino e critério de exclusão ser do sexo masculino.

Exemplo 2: Os critérios de inclusão não dizem respeito à pergunta de pesquisa. Por exemplo, um estudo sobre a função sexual em mulheres puérperas descreve como critérios de inclusão sem história de câncer nos últimos 10 anos. Este, na verdade, corresponde a um critério de exclusão.

Exemplo 3: Não descrição de variáveis-chaves nos critérios de inclusão necessários para interpretar a validade externa. Por exemplo, um estudo em sujeitos com obesidade severa (graus II e III) retrata como critérios de inclusão sujeitos com obesidade, o que incluiu o grau I.

Exemplo 4: Assinatura do Termo de Consentimento Livre e Esclarecido (TCLE). Por exemplo, o estudo elenca como critério de exclusão o não consentimento no estudo. Mas, lembre-se: um participante da pesquisa somente pode ser excluído após ter sido incluído no estudo, ou seja, a assinatura do TCLE é critério de inclusão. O que pode acontecer é a retirada do TCLE ou a perda do seguimento.

O estudo foi bem delineado?

É preciso atentar-se com um dos maiores erros de um pesquisador, o “massacre” dos dados a fim de se obter qualquer resultado. Muitos resultados são distorcidos e interpretados de forma forçada para concluir qualquer coisa. Questiona se o estudo surgiu de uma observação, se foi realizado algum experimento ou intervenção, se há sinais de mineração de dados e se a discussão está de acordo com o método.

Questione: qual foi a intervenção do estudo? Com o que a intervenção foi comparada? O padrão-ouro ou um tratamento de referência foi utilizado? É o comparador adequado? Qual foi a exposição (com o fator)? E a não exposição (sem o fator)? Qual o desfecho medido, e como? Foi utilizado um instrumento validado? Era aplicável as características da população? O método utilizado era sensível e específico? O equipamento estava calibrado?

O viés sistemático foi evitado?

Viés é qualquer fator que influencie erroneamente as conclusões sobre grupos e distorça as comparações. Podem ser relativos às conclusões de uma investigação em relação à amostra estudada (validade interna) ou à extrapolação dessas conclusões para a população de onde a amostra foi retirada ou mesmo para outras populações (validade externa).

A validade externa representa a capacidade de generalização dos resultados de uma pesquisa. Ou seja, a capacidade de os resultados manterem-se verdadeiros em outras situações. A validade externa depende da representatividade da amostra e do grau de veracidade dos resultados de uma observação em outros cenários ou populações. As principais perguntas a se fazer são: as conclusões do estudo podem ser generalizadas para a população de interesse? A amostra representa essa população?

A validade interna representa a solidez da pesquisa quanto a fatores de confundimento e viés. Depende de quão bem a pesquisa é delineada, conduzida e analisada. Em suma, é o grau em que os resultados de um estudo estão corretos para a amostra de pacientes sob análise. Isto é, até que ponto os resultados de um estudo são distorcidos em decorrência de erros metodológicos na concepção do estudo?

Existem dois tipos de erro: aleatório e sistemático. O erro aleatório decorre quando a divergência entre o valor da amostra e o valor verdadeiro (na população) ocorre exclusivamente em decorrência do acaso, ou seja, segue um padrão probabilístico sem ter sua ocorrência interferida por fatores específicos. Esse tipo de erro diminui a precisão (confiabilidade) do estudo e tende a diminuir a associação. Nunca pode ser completamente eliminado

porque quase sempre o estudo é conduzido em uma pequena amostra da população. É suscetível a erro de amostragem, o qual decorre, geralmente, da falta de representatividade da amostra que não contempla toda a variabilidade da população. A melhor forma de reduzir o erro de amostragem é aumentar o tamanho de amostra. Lembre-se: a variação individual sempre ocorre e nenhuma medida é perfeitamente precisa.

O erro sistemático, que corresponde ao viés, vício ou tendenciosidade, expressa a diferença entre a medida real de efeito e aquela obtida no estudo. Está diretamente relacionado à validade interna do estudo. É decorrente de falhas metodológicas ou de condução do estudo. Não pode ser corrigido aumentando o tamanho de amostra. Pode ocorrer devido à existência de vieses e de fatores de confusão. Os dois grandes tipos de vieses incluem o viés de seleção e o viés de informação.

Do ponto de vista estatístico, viés é quando o valor médio da medida de associação obtido a partir de um número infinito de estudos não corresponde ao valor verdadeiro da associação. Do ponto de vista epidemiológico, é o desvio dos resultados ou inferências em relação à verdade, ou o processo que leva a esse desvio. Basicamente, é qualquer tendência na coleta, análise, interpretação, publicação ou revisão de dados que possa conduzir a conclusões sistematicamente diferentes da verdade.

O viés de seleção, como o próprio nome diz, ocorre no momento da seleção/alocação dos participantes da pesquisa. Este viés é evidenciado especialmente quando comparamos os grupos de participantes e notamos que estes diferem em outros determinantes de desfecho (conhecidos ou não), além daquele que está sendo estudado. Ou seja, os grupos de comparação não provêm da mesma população-base do estudo. Pode-se evitar tal viés no planejamento do estudo, no momento em que definimos os critérios de seleção. Os principais tipos de vieses de seleção estão resumidos no Quadro 4.

Quadro 4. Tipos de viés de seleção (adaptado de Bonita e cols., 2010).

Tipo de viés	Definição
Alocação	O pesquisador escolhe um método não-randomizado para distribuir sujeitos nos grupos de estudo.
Amostragem	A amostra não representa o espectro de características da população-alvo. Alguns indivíduos têm maior chance de serem selecionados. Por exemplo, um estudo sobre desnutrição em áreas rurais, onde os investigadores evitam as moradias isoladas e examinam apenas os moradores de regiões próximas à estrada.
Detecção	Resultado da falha em detectar um caso da doença, um possível fator causal ou um efeito de interesse.
Exclusão	Existência de critérios diferentes para a seleção dos participantes que farão parte do grupo de tratamento e para os participantes que comporão o grupo controle.
Hospitalização/ Admissão/ Diagnóstico	Pacientes com determinada característica podem ser mais frequentemente internados que outros sem tal característica (mesma doença). Isso pode levar a conclusão errônea de que essa característica seja um fator de risco, quando na realidade não é. Esse tipo de erro pode causar um exagero sobre o verdadeiro efeito de um fator de risco. Por exemplo, estudo sobre a associação entre Trombose Venosa Profunda (TVP) e uso de contraceptivos orais (CO), as mulheres com sintomas sugestivos de TVP e uso conhecido de CO são mais frequentemente referenciadas ao hospital.
Não-respondentes	É encontrado em estudos transversais, onde os indivíduos não localizados apresentam, tipicamente, maiores prevalências da maioria de doenças do que os que são entrevistados. Em estudos de casos e controles, poderá haver viés se a taxa de não resposta for diferente entre casos e controles, e se a não-resposta estiver associada com a exposição.
Participação	Quando fatores afetam a participação, como exclusões inadvertidas e recusa na participação.
Perdas de Acompanhamento/ Seguimento	Ocorre quando os participantes deixam a pesquisa antes de seu término, encerrando o acompanhamento que era realizado. Caso essas pessoas apresentem maior prevalência da doença do que as pessoas que permaneceram no estudo até o final, teremos o viés de acompanhamento. Por exemplo, os indivíduos que deixam de ser acompanhados podem diferir daqueles que são efetivamente estudados. Isto pode distorcer as medidas de incidência e de efeito, caso as perdas estiverem associadas com a exposição. É comum em estudos de coorte ou intervencionais.
Prevalência/ Incidência	São excluídos os sujeitos curados e falecidos (prevalência) ou os sujeitos com coorte de início (incidência).

Quadro 4. Tipos de viés de seleção (adaptado de Bonita e cols., 2010) (continuação).

Tipo de viés	Definição
Sobrevida Seletiva	Após o diagnóstico, o paciente muda os hábitos associado a doença. Qualquer característica associada com a duração da doença afetará os resultados em um estudo de casos prevalentes (transversais ou caso-controle). Os sobreviventes podem ser atípicos com relação à situação de exposição. Por exemplo, a AIDS é mais frequente em pobres, mas ricos com AIDS sobrevivem mais do que os pobres.
Sobrevivência	Refere-se às situações nas quais você procura pelos erros analisando os acertos.
Trabalhador Saudável/Sadio/Autosseleção/Afiliação	Indivíduos com determinada característica (ligada à doença ou à exposição) têm maior probabilidade de entrar no estudo. Por exemplo, pessoas empregadas são relativamente mais saudáveis do que as desempregadas, aposentadas ou incapacitadas. Provavelmente, menos pessoas empregadas se candidatariam a participar de um estudo.

O viés de informação ocorre por erros na mensuração da exposição e/ou do desfecho no momento da coleta de informações. Ou seja, quando a variável de desfecho é coletada de forma diferente entre os grupos de estudo, ou quando são utilizados diferentes equipamentos ou técnicas de obtenção dos dados. Este tipo de erro é mais facilmente identificado e pode ser corrigido quando existe uma padronização dos equipamentos e instrumentos que serão utilizados na pesquisa ou quando um laboratório central é responsável por todas as análises. Os principais tipos de vieses de informação estão resumidos no Quadro 5.

Quadro 5. Tipos de viés de informação (adaptado de Bonita e cols., 2010).

Tipos de viés	Definição
Aferição/ Mensuração/ Instrumento	Ocorrem quando os métodos de mensuração (instrumentos) empregados diferem entre os grupos de sujeitos ou podem fornecer resultados inadequados para um subgrupo de pacientes. Assim, um grupo tem maior chance de seu desfecho ser detectado do que o outro grupo. Pode ocorrer por diferenças no método, no preparo dos observadores, medidas basais ou dados de seguimento, falta de validade dos instrumentos de coleta de dados, subjetividade de escalas de medidas ou uso de técnicas inadequadas. Por exemplo, aferir a pressão arterial com um aparelho comum em pessoas obesas leva a uma superestimativa dos níveis pressóricos.

Quadro 5. Tipos de viés de informação (adaptado de Bonita e cols., 2010) (continuação).

Tipos de viés	Definição
Aprendizado	Pessoas submetidas ao mesmo instrumento de medida se lembram das respostas corretas e passam a responder conforme o esperado pelo teste e não pelo modo como estão se sentindo.
Detecção/Diagnóstico	Características ligadas à exposição podem aumentar a chance de que uma doença seja detectada. Ocorre quando uma determinada população é monitorada durante um período de tempo. A confirmação da doença poderá ser melhor monitorada nesta população do que na população em geral. Por exemplo, mulheres que fazem uso de anticoncepcionais podem fazer exame citopatológico mais frequentemente do que as demais e, portanto, têm maior probabilidade de detecção de lesões como displasia e carcinoma in situ.
Duração	Quando os casos mais moderados e mais indolentes de uma doença são detectados desproporcionalmente em programas populacionais de rastreamento.
Entrevistador/Observador	O entrevistador coleta diferentemente as informações para doentes e sadios, ou expostos e não exposto, ou ainda direciona a resposta dos participantes. Geralmente, entrevistadores que também participam da interpretação de resultados. Pode ser contornado com o cegamento dos investigadores. Por exemplo, pode ser que o entrevistador saiba que assistir televisão poderá aumentar o consumo de alimentos gordurosos e isso poderá levar ao aumento da obesidade. Ao encontrar uma criança obesa, se esta não relata assistir televisão, o entrevistador insiste, tentando obter uma resposta positiva.
Espectro de Doença	Ocorre quando somente um espectro limitado da doença é incluído.
Exame Tardio	Os casos moderados de uma doença que progridem lentamente são preferencialmente detectados em um inquérito.
Extensão	O espectro completo de um tumor em particular é composto de cânceres que variam de agressivos a indolentes.

Quadro 5. Tipos de viés de informação (adaptado de Bonita e cols., 2010) (continuação).

Tipos de viés	Definição
Publicação	Tendência de que resultados publicados sejam sistematicamente diferentes da realidade. A não publicação dos resultados pode ser por decisão do autor, do financiador do estudo ou de editores por não submeterem ou aceitarem trabalhos com resultados negativos ou desfavoráveis.
Recordação/Memória	Quando a história de exposição é obtida retrospectivamente, os casos podem lembrar-se melhor de sua história de exposição; indivíduos sem a carga da doença, tendem a se esquecer mais facilmente de sua história passada. Em estudos transversais ou de caso-controle, pessoas afetadas por uma doença ou problema podem recordar melhor das exposições no passado do que pessoas não afetadas. Por exemplo, pacientes com câncer de estômago podem se lembrar melhor de certas características da dieta no passado do que indivíduos saudáveis.
Registro	Indivíduos com doenças mais graves tendem a ter registros mais completos sobre exposições e, portanto, pode-se encontrar uma maior associação nestes casos.
Tempo Ganho	O rastreamento detecta uma doença mais precocemente na sua história natural do que teria acontecido de outra maneira.

O fator de confusão ocorre quando a associação observada é devido a um terceiro fator (variável de confusão) que se relaciona com a variável de exposição e com o desfecho, mas não faz parte da cadeia causal existente entre a exposição e o desfecho. Ou seja, ocorre quando dois fatores estão associados e o efeito de um se confunde com o do outro ou é distorcido pelo outro. Em geral, é difícil de ser identificado, comum de ser encontrado e o mais difícil de ser corrigido. Por exemplo, o tabagismo na associação entre consumo de café e desenvolvimento de coronariopatia (Figura 3).

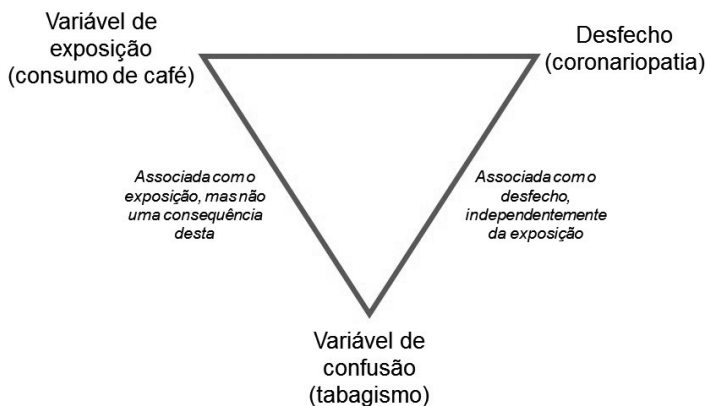


Figura 3. Exemplo de fator de confusão.

Para se evitar um fator de confusão, é importante a sua identificação prévia na elaboração do projeto de pesquisa, a fim de se adaptar o desenho do estudo. Também deve-se adotar métodos de randomização, pareamento, estratificação e restrição de categorias através dos critérios de exclusão. Por isso, deve-se listar os possíveis fatores de confusão, com base na literatura e no modelo teórico, e testar a associação entre cada fator e a exposição. Além disso, pode-se realizar análise multivariável por meio de modelos de regressão (estimar o efeito de cada variável independente sobre a variável dependente).

O estudo teve tamanho e tempo adequados?

Neste último ponto, deve-se observar quatro aspectos importantes:

- 1) o tamanho da amostra;
- 2) a representatividade da amostra;
- 3) a duração do acompanhamento e;
- 4) a totalidade do acompanhamento.

Com relação ao tamanho da amostra, antes de se iniciar qualquer estudo, realiza-se um cálculo para prever o tamanho de amostra necessário para nossa pergunta de pesquisa. Um estudo deve ser suficientemente grande para ter alta chance de detectar, como estatisticamente significativo, um efeito digno de ser notado se ele existir, e assim, prover razoável certeza de que não existe benefício se ele não for encontrado no estudo. Ou seja, o cálculo do tamanho da amostra nos traz informações a respeito da viabilidade do estudo.

Quanto maior a variabilidade de resposta na variável de desfecho e menor a diferença real entre os grupos, maior deve ser o tamanho da amostra para demonstrar se essa diferença realmente existe. Por exemplo, caso sejam estudados dois grupos de gestantes e a prevalência da condição no grupo de estudo é de 40% e no de comparação é de 50%, maior deverá ser o tamanho da amostra. Quanto menor a variabilidade de resposta na variável de desfecho e maior a diferença real entre os resultados dos grupos de estudo, menores devem ser os tamanhos das amostras. Para o cálculo amostral, escolhe-se um nível de significância (α). Esse nível corresponde ao grau de diferença entre os dois grupos que constituiria um efeito clinicamente significativo. Em geral, adotam-se valores de 5% ou 1%. Observe que isso pode não ser o mesmo que um efeito estatisticamente significativo. E, precisamos considerar o que é uma significância estatística dos achados e uma significância clínica.

Resultados clinicamente significantes são aqueles que refletem a cura, a melhora da qualidade de vida, a redução das complicações de uma doença ou a estabilização de um tratamento. Para verificar a existência de significância estatística, devemos informar qual a diferença mínima (α) que deve ser considerada. A diferença encontrada entre os grupos deve ter tamanho suficiente para ter importância para a prática clínica. Além disso, a amostra deve ter boa representatividade com perfil que transpareça suas semelhanças e diferenças do universo. É preciso analisar o perfil da população, a fim de identificar o quanto essa população representará o universo.

Conforme a natureza do desfecho de interesse, para o cálculo, também é utilizada a medida de tendência central, a média, e de dispersão, o desvio padrão. Se o desfecho em questão for um evento (como uma histerectomia) e não uma quantidade (como a medida da pressão arterial), os itens dos dados requeridos são a proporção (prevalência, incidência) de pessoas sofrendo o evento na população e uma estimativa do que constituiria uma alteração clinicamente significativa nesta proporção.

Antes de iniciar o estudo, os pesquisadores podem descobrir que tamanho deve ter uma amostra para que tenham uma chance moderada, alta ou muito alta de detectar uma diferença verdadeira entre os grupos. A probabilidade de detectar uma diferença verdadeira é conhecida como poder do estudo. É comum que os estudos estipulem um poder entre 80 e 90%.

Importante: Quando há mais de um desfecho de interesse, o tamanho da amostra deve ser calculado para cada um deles, adotando-se o maior valor. Se houver modificação no delineamento de pesquisa, será necessário realizar um novo cálculo amostral.

Mas, de onde extraímos esses dados? Buscaremos esses valores em artigos científicos que estudaram desfechos semelhantes, com populações ou subgrupos similares quanto a características de doença, faixa etária, massa corporal, sexo, etc. Caso não haja estudos disponíveis sobre o tema para que se possa extrair essas informações, é recomendado a realização de um estudo piloto.

Existem diversos softwares e websites disponíveis para o cálculo de tamanho amostral, como: MACORR (<http://macorr.com/sample-size-calculator.htm>), PASS (<https://www.ncss.com/software/pass/>), Sealed Envelope (<https://www.sealedenvelope.com/power/>) e Winpepi (<http://www.brixtonhealth.com/pepi4windows.html>).

E se não houver o cálculo do tamanho amostral? Se o artigo que você estiver lendo não fornece um cálculo do tamanho da amostra e parece mostrar que não há diferença entre os grupos, você deve extrair do artigo as informações do α , média e desvio

padrão e calcular o tamanho de efeito da amostra e estimar seu poder. Apenas o valor de p ser significativo não garante que o efeito é significativo ou importante. Estudos com baixo poder ocorrem geralmente porque os pesquisadores tiveram dificuldade em recrutar os participantes da pesquisa ou porque não haviam pré-determinado um tamanho amostral mínimo necessário para encontrar a significância esperada.

Da mesma forma, para determinar o tempo de acompanhamento de um estudo ou a duração de um experimento, utilizam-se como base informações de artigos científicos ou protocolos de pesquisa. Nos estudos com acompanhamento, há algumas razões pelas quais os participantes da pesquisa acabam retirando o consentimento ou são desligados pelos pesquisadores, como: inclusão incorreta do participante no estudo; suspeita de reação adversa ao medicamento ou tratamento; perda de motivação do paciente; razões clínicas (por exemplo, doença concomitante, gestação); perda de acompanhamento (por exemplo, o participante muda de endereço, de telefone) e; morte. Por isso, devem ser incluídas as perdas no cálculo do tamanho de amostra, que são estimadas, em média, entre 10 e 20%.

Resumindo, ao observar o método de um estudo, é possível chegar às seguintes informações: tipo de estudo, número de participantes, origem dos participantes, tratamento ou intervenção, exposição ou não exposição, tempo de acompanhamento, medidas de desfecho, análise estatística e cálculo do tamanho amostral. Fazendo isso, será mais fácil compreender, interpretar e, se apropriado, rejeitar os resultados do estudo.

Referências

Bonita R, Beaglehole R, Kjellström T. Epidemiologia básica. 2. ed. São Paulo: Santos, 2010. 213 p.

Elsevier. Ethics in Research & Publication, 2019. Disponível em https://www.elsevier.com/__data/assets/pdf_file/0012/856659/Ethics-in-Research-and-Publication-March-2019.pdf

Greenhalgh T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5a ed. Porto Alegre: Artmed, 2015. 376 p. Moreira de Sá RA, Oliveira MA, Vellarde GC. Entendendo a pesquisa clínica VI: formatando a pergunta do estudo. FEMINA. 2015;43(6):277-280.

Van Bavel JJ. Struggling with your academic writing? Try these experiments to get the words flowing. Science. 2019. Disponível em: <https://www.sciencemag.org/careers/2019/06/struggling-your-academic-writing-try-these-experiments-get-words-flowing>. Acesso em 25/08/2020.

Estudos secundários: revisão sistemática e metanálise

*Fernanda Dapper Machado
Fernanda Vargas Ferreira
Edison Capp
Otto Henrique Nienov*

Como saber se um medicamento é eficaz e seguro para o tratamento de uma doença? Para responder essa pergunta de pesquisa é indicado a condução de um ensaio clínico randomizado. Mas, um único ensaio clínico randomizado seria suficiente para responder a essa pergunta e modificar alguma conduta? Será essa pergunta original, ou seja, nenhum outro estudo propôs respondê-la anteriormente? Para responder essas perguntas existem as revisões sistemáticas e metanálises, que possibilitam comparar as evidências disponíveis, avaliar rigorosamente os estudos originais (se estes foram ou não bem conduzidos) e, ao final, gerar uma recomendação para tomada de decisão.

O rigor metodológico é essencial na execução de qualquer estudo epidemiológico, especialmente nas revisões sistemáticas. Para uma melhor avaliação do rigor metodológico existem guias (*checklists*) que auxiliam na avaliação dos estudos incluídos em uma revisão sistemática. Esses guias metodológicos são úteis e indicados para a leitura crítica de qualquer artigo científico.

A evidência gerada por um estudo de revisão sistemática permite a criação de “pontes” entre o mundo ideal e o real, auxiliando na tomada de decisão no contexto da Medicina Baseada em Evidências. Desta forma, as revisões sistemáticas são consideradas o tipo de estudo com maior nível de evidência, posicionados no topo da pirâmide do nível das evidências, utilizadas para subsidiar a tomada de decisão.

Quando uma revisão é sistemática?

A revisão é um processo de busca, análise e descrição de um corpo do conhecimento em busca de resposta a uma pergunta. Existem diferentes tipos de revisão: narrativa, integrativa, sistemática, de escopo e de mapeamento, definidas de acordo com o método de elaboração. Mas, quando uma revisão é de fato sistemática?

A revisão narrativa ou tradicional é uma revisão mais ampla, cujo processo abrangente não utiliza critérios específicos, explícitos e sistemáticos de busca, análise e descrição de um corpo de evidências em busca de resposta a uma pergunta de pesquisa ampla. Por não utilizar critérios de inclusão e de exclusão estabelecidos e sistemáticos para busca e análise crítica dos estudos, é chamada de *black box*. Desta forma, a revisão narrativa não se propõe a esgotar a busca por todas as informações disponíveis, bem como não aplica estratégias de busca sofisticadas e exaustivas como na revisão sistemática. Geralmente subsidiará fundamentação teórica para outros trabalhos acadêmicos (como teses e dissertações). Outrossim, aplica-se para estruturação conceitual e contextualização do tema, não necessitando de protocolo com informações sobre a seleção dos estudos e a interpretação das informações, as quais podem expressar a subjetividade dos autores.

A revisão integrativa surgiu como uma alternativa para revisar rigorosamente e combinar estudos com diversos delineamentos, não se limitando a um único desenho de estudo, mas, por sua vez, mantendo o rigor metodológico das revisões sistemáticas. O objetivo principal é obter um profundo entendimento de um determinado tema baseando-se em estudos anteriores. Permite a inclusão de estudos experimentais e observacionais, integrando seus resultados, a fim de definir um cenário mais amplo. Agrega conceitos, literatura empírica e teórica, reunindo todas as evidências para definir conceitos, identificar lacunas, revisar teorias e analisar metodologicamente os estudos sobre um determinado tópico. Como a revisão narrativa permite a inclusão de múltiplos estudos com diferentes delineamentos de pesquisas, a realização de análises como a metanálise fica limitada.

Frequentemente utilizada na área da Enfermagem, a revisão integrativa envolve seis etapas:

- 1) definição de um problema norteador e formulação de uma hipótese ou questão de pesquisa relevante, clara, específica e interessante;
- 2) estabelecimento de critérios para inclusão e exclusão de estudos e busca nas bases de dados;
- 3) definição, organização e sumarização das informações a serem extraídas dos estudos selecionados;
- 4) avaliação crítica dos estudos incluídos;
- 5) interpretação e discussão dos resultados e;
- 6) elaboração do documento com a divulgação dos resultados da revisão.

A revisão sistemática tem como objetivo levantar, reunir, avaliar criticamente a metodologia da pesquisa e sintetizar os resultados de estudos primários. Utiliza métodos sistemáticos e explícitos para recuperar, selecionar e avaliar os estudos incluídos na revisão, o que permite a sua reprodutibilidade. Reúne evidências para responder uma pergunta de pesquisa claramente definida e delimitada, com critérios de inclusão e exclusão bem estabelecidos. As fontes de busca são abrangentes e em múltiplas bases de dados, incluindo a literatura cinzenta. As inferências são realizadas conforme todas as evidências disponíveis, permitindo a elaboração de recomendações. A metanálise corresponde à síntese estatística das evidências encontradas em uma revisão sistemática, embora nem toda revisão sistemática permita a execução de uma metanálise.

A revisão de escopo é considerada precursora da revisão sistemática e se limita a mapear os principais conceitos de determinada área do conhecimento, examinando-se sua extensão, alcance e natureza, sumarizando os dados da investigação e identificando possíveis lacunas ou *gaps* na literatura. Por si só, não agrega evidência científica. Por fim, a revisão de mapeamento estabelece novos conceitos, conexões e relações. É útil para reorganizar o conhecimento, sendo amplamente utilizada em *brainstorming*, colaboração, entendimento e inovação de um tema.

Revisão sistemática

A revisão sistemática é uma técnica científica objetiva, transparente, imparcial, eficiente e reprodutível, que permite extrapolar achados de estudos independentes, avaliar a consistência de cada um deles e explicar as possíveis inconsistências e conflitos. A estratégia de busca, dado o rigor metodológico necessário, deve ser definida com o estabelecimento de critérios de inclusão e exclusão dos estudos, especialmente, porque estes constituem os sujeitos (amostra) da pesquisa. Adicionalmente, tais critérios devem ser justificados a fim de demonstrar o caráter de elegibilidade.

O processo de desenvolvimento da revisão sistemática inclui a caracterização de cada estudo selecionado, avaliação da qualidade metodológica, identificação de conceitos importantes, comparação das análises estatísticas apresentadas e conclusão sobre a informação encontrada. O trabalho de revisão sistemática segue a estrutura de um artigo original. A intenção é que a revisão possa ser reproduzida por outro grupo e que o mesmo resultado seja encontrado.

A revisão sistemática é considerada um estudo secundário que reúne e sumariza as evidências de diferentes estudos que respondem à uma mesma questão de pesquisa. É possível realizar revisões sistemáticas de ensaios clínicos randomizados, ensaios clínicos não randomizados, estudos observacionais, testes de diagnóstico ou de rastreamento, dados de pacientes individuais e também revisão de revisões (*overview*).

Apesar de ser considerada um estudo secundário, a revisão sistemática é um estudo original pois responde a uma pergunta de pesquisa específica e relevante e, quando acompanhada de metanálise, fornece mais poder e precisão ao tamanho do efeito. Normalmente os estudos primários não conseguem detectar ou excluir a diferença entre intervenções. Nesse sentido, a revisão sistemática fornece uma evidência nova, portanto, original, com maior grau de confiabilidade.

A metanálise, que se conceitua como tratamento estatístico das evidências, permite aumentar a acurácia dos resultados, melhorando a precisão das estimativas de efeito de uma determinada intervenção clínica. Dessa forma, conhecimentos

de Bioestatística e Epidemiologia são essenciais no que tange à escolha do delineamento para responder à pergunta e para a execução da metanálise.

Portanto, a revisão sistemática é um processo sistemático dos resultados dos estudos, que envolve desde a definição dos objetivos e dos critérios de elegibilidade da revisão, a busca pelos estudos que atendem a esses critérios, a tabulação da forma mais completa possível das evidências de cada estudo identificado, a aplicação dos critérios de elegibilidade e justificativa de quaisquer exclusões, e a análise da qualidade metodológica, de preferência com o auxílio de investigadores com expertise na área da revisão. Por fim, analisa os resultados dos estudos elegíveis, utilizando síntese estatística (metanálise), se apropriada e possível, e compara as análises alternativas, preparando um resumo crítico da revisão, definindo objetivos, descrevendo material e métodos e relatando os resultados.

Os pesquisadores devem se responsabilizar pela atualização das revisões sistemáticas, atualizando a busca através da inclusão de novos estudos. Para isso, recomenda-se a realização de buscas exploratórias periódicas (mensais ou semestrais) após a publicação da revisão para verificar a necessidade de atualização.

Por que fazer uma revisão sistemática?

Na maioria das perguntas de pesquisa, um único estudo não é suficiente para que se tenha confiança na evidência devido à quantidade de informações científicas disponíveis e, portanto, é imprescindível que as informações sejam reunidas, organizadas, criticamente avaliadas e quantitativamente mensuradas.

A revisão sistemática é uma etapa importante na elaboração de diretrizes clínicas e análises econômicas, permitindo uma ligação entre a evidência científica e a prática clínica, sendo extremamente útil para os tomadores de decisão. Gestores e formuladores de políticas devem utilizar a melhor evidência disponível para o processo de tomada de decisão, considerando que essa definição impactará a vida da sociedade e auxiliará na

alocação de recursos.

Quanto tempo leva uma revisão sistemática?

Um primeiro aspecto a ser considerado é de que uma revisão sistemática não pode ser feita sozinha. É um trabalho que envolve a participação de uma equipe de especialistas, que irão auxiliar a esclarecer questões relacionadas ao tema da revisão, como bibliotecários para desenvolver estratégias de pesquisa abrangentes e identificar bancos de dados apropriados; revisores para selecionar resumos e realizar a leitura dos textos completos de forma independente; estatísticos para ajudar na análise de dados e; um líder, especialista no tema do projeto, para coordenar a revisão sistemática.

Se há a necessidade de fazer uma revisão sistemática em um semestre, talvez não seja o tipo de pesquisa adequada para este prazo. As etapas previstas para uma revisão sistemática, sem considerar potenciais dificuldades, como perda de dados, não retorno do contato com os autores e não aceite da publicação, demandam entre 6 a 18 meses para elaboração, mesmo por grupos mais experientes. O tempo de elaboração de uma revisão sistemática se assemelha a de uma pesquisa primária, em que são consumidos tempo para coleta, análise, redação e publicação dos resultados. É um investimento de tempo considerável que requer que os pesquisadores disponham de tempo e recursos para a sua elaboração. O quadro 1 descreve o tempo estimado para a execução de cada uma das etapas da revisão sistemática, sendo que algumas destas ocorrem concomitantemente.

Os recursos financeiros de uma revisão sistemática incluem recursos humanos (revisores, coordenador do projeto, bibliotecários, especialistas ou estatísticos) e materiais, (computadores, materiais de escritório, softwares de gerenciamento bibliográfico e estatístico).

Quadro 1. Tempo estimado para execução de cada uma das etapas da revisão sistemática. Fonte: Adaptado de Cochrane Library.

Tempo estimado (em meses)	Atividade
1 - 2	Elaboração do protocolo
3 - 8	Busca nas bases de dados
2 - 3	Teste piloto de critérios de elegibilidade
3 - 8	Seleção de estudos
3	Teste piloto para avaliação de risco de viés
3 - 10	Avaliação de validade
3	Teste piloto de coleta de dados
3 - 10	Coleta de dados
3 - 10	Imputação de dados
5 - 11	Acompanhamento de informações ausentes
8 - 10	Análise
1 - 11	Preparação do manuscrito
12	Atualização da revisão sistemática

Vantagens da revisão sistemática

Devido ao rigor metodológico, a elaboração de um protocolo da revisão sistemática limitará o risco de viés, especificamente, para identificar estudos não condizentes com a proposta. A análise dos dados pode ser interpretada com facilidade, mediante comparações viáveis entre distintos estudos a fim de se definir a generalidade dos achados e a consistência (falta de heterogeneidade) dos resultados. Tal premissa gerará maior confiabilidade e precisão no conjunto de evidências.

Estudos menores, que são a norma em muitas áreas, podem demonstrar uma tendência de impacto positivo, porém podem ser limitados pela ausência de significância estatística. Quando dados de diversos estudos pequenos são somados matematicamente em uma metanálise, os dados combinados podem produzir um achado estatisticamente significativo com provável plausibilidade biológica e inserção na prática clínica.

Revisões sistemáticas podem ajudar a resolver achados contraditórios entre estudos diferentes sobre a mesma questão. Além disso, a heterogeneidade (inconsistência entre resultados de estudos) pode ser identificada e novas hipóteses podem ser geradas sobre subgrupos específicos. Ou seja, a revisão sistemática pode inclusive levantar novas perguntas de pesquisa, propiciando resultados robustos e passíveis de generalização, especificamente, mediante rigor metodológico. Por fim, o tempo entre as descobertas de pesquisa e a implementação das estratégias diagnósticas e terapêuticas efetivas é reduzido quando temos uma revisão sistemática bem delineada disponível.

Limitações da revisão sistemática

Por se tratarem de estudos secundários, as revisões sistemáticas podem reproduzir e ampliar as falhas dos estudos originais. Por exemplo, se todos os estudos primários consideraram um medicamento em dose subterapêutica, a conclusão geral e possivelmente equivocada pode ser de que tal medicamento não apresenta efeito. Por resumir toda a evidência disponível, e considerando as peculiaridades de cada um dos estudos primários, a leitura de uma revisão sistemática pode ser complexa.

Como conduzir e avaliar uma revisão sistemática?

Para que uma pesquisa seja considerada revisão sistemática, as seguintes etapas básicas devem ser realizadas:

- 1) formulação da pergunta de pesquisa;
- 2) estratégia de busca;
- 3) busca na literatura;
- 4) seleção dos estudos;
- 5) extração dos dados;
- 6) avaliação da qualidade metodológica dos estudos incluídos;
- 7) síntese dos dados;
- 8) avaliação da qualidade das evidências;
- 9) redação e publicação dos resultados, e;
- 10) atualização da revisão sistemática.

Etapa 1: Formulação da pergunta de pesquisa

Nesta primeira etapa inclui-se: definição da pergunta de pesquisa, elaboração e planejamento do projeto, contato com pesquisadores, composição da equipe, realização de estudo piloto das buscas, registro do protocolo e submissão às agências de fomento.

Definição da pergunta de pesquisa

O primeiro passo de uma revisão sistemática é definir uma pergunta de pesquisa que seja factível de ser respondida. Ao contrário das revisões narrativas, revisões sistemáticas devem responder a uma pergunta bem estruturada e com um foco bem definido. Para nortear a elaboração da pergunta de pesquisa, recomenda-se o uso do acrônimo PICOS: população ou problema, intervenção ou exposição, comparação ou não-exposição, *outcomes* (desfechos) e tipo de estudo (delineamento que melhor responde à pergunta de pesquisa) (Figura 1).

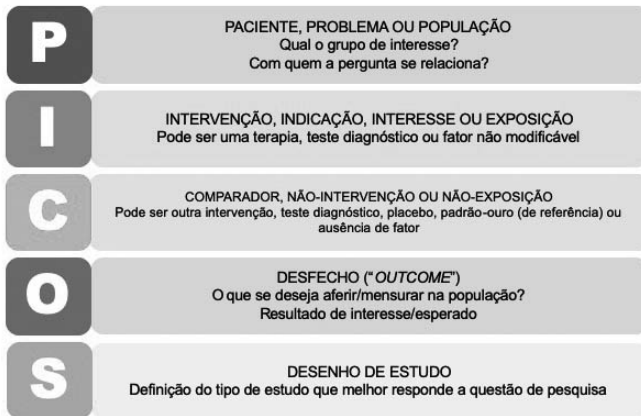


Figura 1. Acrônimo PICOS.

Da mesma forma que a pergunta de pesquisa precisa ser bem delimitada, ao realizar a leitura crítica de uma revisão sistemática, deve-se questionar: qual é a importante questão de pesquisa que a revisão abordou? A questão abordada deve ser definida com muita precisão, pois o(s) revisor(es) deve(m) tomar uma decisão dicotômica, a favor ou contra a inclusão do artigo.

A caracterização do desfecho é uma parte crucial da elaboração da PICOS. Neste momento o pesquisador deve-se perguntar: qual o desfecho mais importante para a pergunta de pesquisa? Qual o conjunto mínimo de desfechos que devem ser medidos e relatados em todos os ensaios clínicos relacionados a uma doença específica ou população de estudo? Há uma iniciativa que buscou desenvolver uma diretriz sobre como selecionar instrumento de medição de resultados, a *COMET initiative* (www.comet-initiative.org/). É possível fazer a busca por desfechos de acordo com a doença de interesse.

É importante revisar também se a pergunta de pesquisa é factível, interessante, inovadora, ética e relevante (acrônimo FINER) (Figura 2).

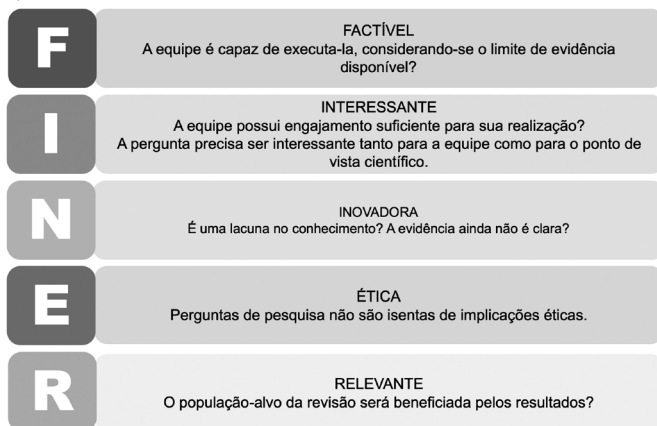


Figura 2. Acrônimo FINER.

Em alguns casos precisamos revisar sistematicamente a literatura para verificar qual a ocorrência de um fenômeno em determinada localidade. Não se trata de observar os efeitos, mas conhecer a dimensão do problema. Trata-se de revisão sistemática de frequência. Neste caso a pergunta de pesquisa deve especificar a população e o fenômeno de interesse, além de definir o delineamento dos estudos (acrônimo POT) (Figura 3).

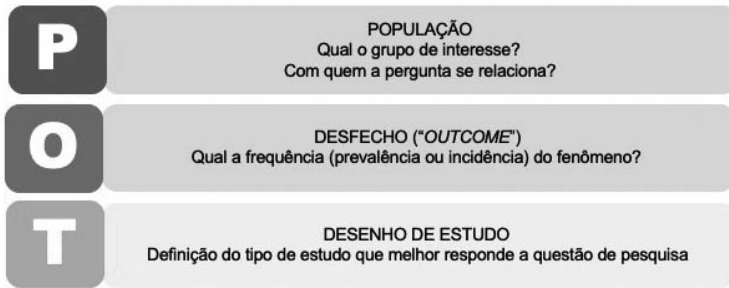


Figura 3. Acrônimo POT.

Este tipo de revisão sistemática é utilizado para análise de impacto orçamentário e estimativa de carga de doenças baseada na estimativa dos anos de vida ajustados por incapacidade (*disability-adjusted life year, DALY*). O DALY é uma medida de carga global de doença. Mede-se simultaneamente os efeitos da mortalidade e os problemas de saúde que afetam a qualidade de vida dos indivíduos. Um DALY é equivalente a um ano de vida sadia perdido. Incorpora os conceitos de mortalidade medida pelos anos de vida perdidos por morte prematura (*years of life lost, YLL*) e de morbidade medida pelos anos vividos com incapacidade (*years lived with disability, YLD*).

É possível fazer revisões sistemáticas de prevalência ou incidência de doenças. Lembrando que a prevalência se refere a indivíduos com a doença ou agravo em um período de tempo, ou seja, uma fotografia da situação que inclui tanto casos novos como casos antigos. A incidência, por sua vez, contabiliza os casos novos em determinado período, sendo equivalente a um filme.

Também é possível realizar revisão sistemática de pesquisa qualitativa. Neste caso, para a pergunta de pesquisa, utilizamos o acrônimo PICO: paciente, problema ou população, interesse e contexto (Figura 4). Por exemplo, qual a percepção da parturiente (paciente) com a presença de acompanhante (interesse) na sala de pré-parto (contexto)?

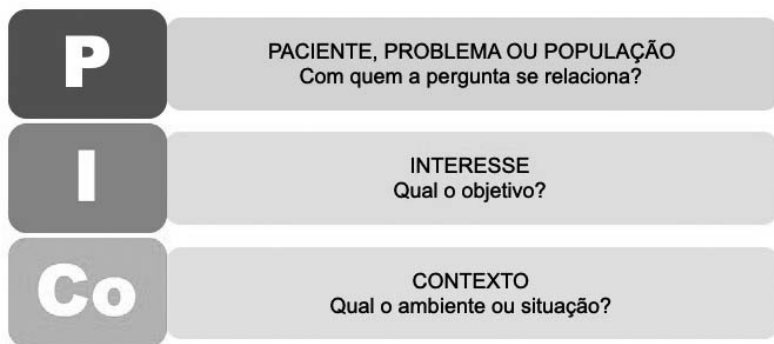


Figura 4. Acrônimo PICo.

Importante! Compartilhar a pergunta de pesquisa da revisão sistemática com algum colega, com a finalidade de aprimorá-la, levando em consideração os acrônimos PICOS e FINER. Além da pergunta de pesquisa, os critérios de elegibilidade também precisam ser definidos a priori e registrados no protocolo da revisão sistemática. Os critérios de inclusão e de exclusão complementam a pergunta de pesquisa, refinando-a.

Protocolo da revisão sistemática

Elaborar um protocolo da revisão sistemática é essencial, mas, antes da elaboração do protocolo e tendo definida a pergunta de pesquisa, é recomendado verificar as revisões e protocolos existentes nas bases de dados e de registro. O registro do protocolo aumenta a transparência e permite a reprodutibilidade da pesquisa, evitando a duplicação de esforços sobre um mesmo tema. O protocolo de uma revisão sistemática é muito semelhante a um projeto de pesquisa original: contempla uma pergunta de pesquisa, traduzida em objetivo. Para respondê-la, há um método detalhado e factível de ser executado.

O protocolo da revisão sistemática deverá apresentar a pergunta de pesquisa, a justificativa para a sua elaboração e os objetivos. A metodologia deverá conter os critérios de inclusão (ou seja, colocar a questão de pesquisa em prática e definir população, intervenção ou comparação, resultado e desenho

de estudo), critérios de exclusão (para não incluir determinados estudos e populações específicas), critérios de elegibilidade (qualitativos e quantitativos ou metanálise), estratégia de pesquisa (deve ser explícita, reprodutível e incluir literatura cinzenta), avaliação crítica (análise do risco de viés nos estudos individuais) e avaliação da qualidade metodológica (evidência gerada pela revisão sistemática). Por fim, a síntese dos resultados (interpretação, análise e aplicabilidade dos resultados).

Guias metodológicos

Alguns guias metodológicos oferecem suporte para a redação e execução de revisões sistemáticas. Entre eles destacam-se o Manual da Cochrane (training.cochrane.org/handbook), o Manual do Centro de Revisões da Universidade de York (www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/SysRev3.htm) e o Guia de Redação de Protocolos de Revisão Sistemática (PRISMA) (www.prisma-statement.org/). Além dos guias disponíveis para a elaboração do protocolo da revisão sistemática, é fundamental o conhecimento da equipe, ou de pelo menos um dos membros, sobre o tema da pesquisa.

O PRISMA-P (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols*) corresponde a uma lista de verificação estruturada para desenvolver uma revisão sistemática e metanálise. Também serve como uma excelente base para a leitura de uma revisão sistemática. O PRISMA faz parte da Equator Network (www.equator-network.org/), que desenvolve e disponibiliza roteiros para a redação de artigos científicos para melhorar a confiabilidade e transparência da pesquisa (Quadro 2).

É indicado realizar o registro do protocolo da revisão sistemática quando ele estiver suficientemente maduro, a fim de garantir todas as condições para a sua realização. Idealmente, devem ser realizados pilotos da busca, seleção e testes da extração com algum dos estudos potencialmente elegíveis para melhor planejar e descrever as etapas que serão realizadas. Registrar o protocolo sem ter as condições para executá-lo, além de desperdício de tempo, pode favorecer a execução da revisão sistemática por outros grupos. A equipe de revisores deve ter experiência e tempo para executá-la

de acordo com todos os critérios que garantam a sua qualidade.
 Quadro 2. Guias para redação de artigos científicos de acordo com o tipo de estudo.

Tipo de estudo	Guia para redação
Diretrizes clínicas	<i>Appraisal of Guidelines for Research and Evaluation (AGREE)</i> (www.agreetrust.org/)
Ensaio clínico randomizado	<i>Consolidated Standards of Reporting Trials (CONSORT)</i> (www.consort-statement.org/)
Estudos de análise econômica	<i>Consolidated Health Economic Evaluation Reporting Standards (CHEERS)</i> (DOI: 10.1136/bmj.f1049)
Estudos diagnósticos	<i>Standards for Reporting Diagnostic Accuracy (STARD)</i> (DOI: 10.1136/bmj.h5527)
Estudos em modelo animal	<i>Animal Research: Reporting of In Vivo Experiments (ARRIVE)</i> (arriveguidelines.org/)
Estudos observacionais	<i>Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)</i> (www.strobe-statement.org/)
Estudos qualitativos	<i>Standards for Reporting Qualitative Research (SRQR)</i> (DOI: 10.1097/ACM.0000000000000388) e <i>Consolidated criteria for reporting qualitative research (COREQ)</i> (DOI: 10.1093/intqhc/mzm042)
Modelos de predição prognóstica	<i>Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)</i> (www.tripod-statement.org/resources/)
Protocolo de ensaio clínico	<i>Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)</i> (www.spirit-statement.org/)
Protocolo de revisão sistemática	<i>Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P)</i> (DOI: 10.1186/2046-4053-4-1)
Relato de caso	<i>Consensus-based Clinical Case Reporting Guideline Development (CARE)</i> (www.care-statement.org/)
Revisão sistemática	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)</i> (www.prisma-statement.org/)

O protocolo pode ser registrado em periódicos científicos que aceitam esse tipo de publicação. Geralmente são revistas com políticas *Open Access*, para as quais os autores pagam taxa de publicação após o aceite do artigo e este fica disponível gratuitamente aos leitores.

A iniciativa PROSPERO (*International Prospective Register of Systematic Reviews*) (www.crd.york.ac.uk/PROSPERO/), do Centre for Reviews and Dissemination (CRD) da Universidade de York, de uso prioritário para revisões do Reino Unido, permite o registro gratuito e aberto a qualquer pessoa que realize revisões sistemáticas dos efeitos de intervenções e estratégias para prevenir, diagnosticar, tratar e monitorar as condições de saúde, para as quais há um resultado relacionado à saúde. Lembre-se que o registro garante transparência e reprodutibilidade. Na mesma base de registro é possível realizar a busca por outros registros de protocolos e, assim, evitar duplicidades. Além do registro de revisões sistemáticas de estudos em que os participantes são pessoas, na PROSPERO é possível realizar o registro de revisões sistemáticas em que os sujeitos dos estudos são animais.

Outros registros disponíveis são o *Open Science Framework* (osf.io/), do *Centre for Open Science*, um repositório aberto para projetos de todas as áreas; a *Campbell Systematic Reviews* (onlinelibrary.wiley.com), da *Campbell Collaboration*, de revisões elaboradas pela colaboração e; a *Cochrane Library* (www.cochranelibrary.com), da *Cochrane*, também de revisões elaboradas pela colaboração.

A revisão sistemática necessita de recursos para a pesquisa?

Da mesma forma que uma pesquisa de um estudo primário necessita de recursos para sua execução, a pesquisa de uma revisão sistemática também necessita viabilizar os recursos para o bom andamento do planejamento previsto no protocolo. Alguns aspectos merecem atenção: espaço físico e recursos materiais, informacionais e humanos.

A equipe tem espaço para trabalhar? O ideal é que a equipe tenha disponível um local físico para desenvolver suas atividades e fazer reuniões de acompanhamento das etapas da revisão sistemática. Neste aspecto, a ergonomia também deve ser priorizada, já que essa permite a ligação ser humano-trabalho e pode influenciar sobre a produtividade.

Os recursos materiais estão disponíveis na quantidade e qualidade necessárias? Computadores com acesso à internet, de processamento rápido e eficaz, impressoras e outros insumos devem estar disponíveis à toda equipe.

Os recursos informacionais estão disponíveis? Softwares de referência, de gestão da revisão, de análise estatística e pacotes de armazenamento em nuvem permitem a execução do trabalho de forma segura e ágil. Além disso, permitem o acesso de toda a equipe em qualquer local.

Os recursos humanos reúnem as expertises básicas? A equipe deve ter domínio das tecnologias, organização dos arquivos e processos, conhecimentos básicos de epidemiologia e estatística, além daqueles específicos de revisão sistemática. Pelo menos um ou mais membros da equipe ou um especialista facilmente acessível deve ter conhecimento sobre o tema da revisão para resolução de dúvidas, tanto clínicas como metodológicas.

Etapa 2: Estratégia de busca na literatura

Antes de iniciar esta etapa é importante se certificar de que a pergunta de pesquisa da revisão sistemática está adequada, clara e definida. Se a pergunta de pesquisa estiver mal formulada ou não for específica, provavelmente a estratégia de busca será inadequada e/ou mal conduzida, o que levará à síntese incorreta dos resultados e não se obterá a evidência necessária para responder o objetivo da revisão sistemática.

Ao ler uma revisão sistemática, é importante perguntar se: foi realizada uma pesquisa detalhada nas bases de dados apropriadas e foram exploradas outras fontes potencialmente importantes? Em uma revisão sistemática, diferentemente de uma revisão narrativa, é imprescindível identificar a fonte da informação e como ela foi processada.

A formulação da estratégia de busca na literatura engloba sua elaboração e testagem, eleição das fontes para a busca bibliográfica, buscas nas bases de dados, exportação dos resultados, buscas manuais (por exemplo, rastreamento de referências de estudos) e buscas em fontes de literatura cinzenta (por exemplo, anais de congressos não indexados, repositórios de teses e dissertações).

Primeiro aspecto sobre a busca é de que não existe estratégia de busca rápida, pois a medida que realizamos a busca e leitura de artigos, vamos encontrando novos artigos, novas nomenclaturas, novos termos e, vamos aperfeiçoando a estratégia de busca. Por isso, essa etapa da revisão sistemática denota tempo, pois a busca precisa ser abrangente.

Escolha das bases de dados

A escolha das bases de dados utilizadas na busca deve levar em conta o que já planejamos para a revisão sistemática, a pergunta de pesquisa e os critérios de elegibilidade. O que você quer encontrar indica onde você deve procurar. A Diretriz Metodológica para elaboração de revisões sistemáticas e metanálises elaborada pelo Ministério da Saúde classifica como essencial a pesquisa nas bases CENTRAL, MEDLINE, EMBASE e LILACS, considerando que a CENTRAL reúne registros de ensaios clínicos, a MEDLINE tem ênfase na literatura científica dos Estados Unidos, a EMBASE tem ênfase na produção europeia (com apenas 34% de sobreposição à MEDLINE) e o LILACS enfatiza a produção científica latino-americana.

Ao realizar uma revisão sistemática, deve-se não somente buscar por artigos relevantes de forma cuidadosa e objetiva, mas também explicitar os critérios usados para rejeitar artigos como sendo “falhos”, independentemente dos resultados que apresentem. Em outras palavras, não se pode desprezar um estudo porque todos os outros estudos mostraram algo diferente. A exclusão deve ocorrer porque, independentemente dos resultados encontrados, os objetivos ou métodos do estudo não atenderam aos critérios de inclusão ou de padrão de qualidade propostos para a revisão.

A maior parte das bases de dados depende de assinatura, que geralmente é viabilizada por acesso institucional das Universidades ou pelo Portal de Periódicos Capes (Quadro 3). Como isso depende de contratos, períodos de descontinuidade podem ocorrer e devem ser considerados na escolha das bases de dados, pois o acesso precisa ser garantido até a revisão ser concluída. Além disso, é necessário que a busca seja atualizada enquanto a revisão estiver sendo elaborada até o momento em que for concluída e publicada e, em seguida, regularmente, para a avaliação de possíveis atualizações.

Elaboração da estratégia de busca

A estratégia de busca utilizada em uma revisão sistemática deve ser a mais sensível possível. Para isso, é importante diferenciarmos os tipos de busca: simples, avançada e sensível. A busca simples pode ser definida como uma busca genérica, como aquelas feitas no Google ou uma busca simples no PubMed. A busca avançada é uma busca mais específica, como aquela realizada na busca avançada do PubMed utilizando os MeSH terms e Entry terms. Já a busca sensível corresponde a uma busca aprimorada, onde além de incluir os MeSH terms e Entry terms no PubMed, deve-se incluir sinônimos e realizar uma busca utilizando estes termos também como *Text terms*.

Portanto, a construção da estratégia de busca da revisão sistemática prioriza a sensibilidade em detrimento à especificidade, gerando um número muito maior de estudos do que as que realmente preenchem os critérios de elegibilidade. A busca nas bases de dados é um processo sofisticado, e mesmo a melhor busca corre o risco de deixar artigos potencialmente importantes de fora.

Cada base de dados possui um vocabulário controlado ou thesaurus (palavras com significados semelhantes, dentro de um domínio específico de conhecimento). Por exemplo, o PubMed tem como *thesaurus* os *MeSH terms*, enquanto que a Biblioteca Virtual em Saúde utiliza os descritores em saúde (Decs). A elaboração da estratégia de busca deve considerar grafias alternativas, terminações alternativas, sinônimos, nomes comerciais ou genéricos de medicamentos, antônimos e acrônimos. Há também especificidades

Quadro 3. Bases de dados utilizadas para busca.

Base de dados	Área de Conhecimento	Acesso
<i>AMED (Allied and Complementary Medicine)</i>	Medicina complementar e tratamentos alternativos	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>CENTRAL (The Cochrane Register of Controlled Trials)</i>	Ciências da Saúde	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>CINAHL (Cumulative Index to Nursing and Allied Health Literature)</i>	Enfermagem e terapia ocupacional	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>ECRI</i>	Equipamentos biomédicos	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>EMBASE (Excerpta Medica Database)</i>	Ciências da Saúde	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>ERIC (Education Resources Information Center)</i>	Educação em Saúde	Aberto e gratuito
<i>IPA (International Pharmaceutical Abstracts)</i>	Farmacologia	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>LILACS (Literatura Latino-Americana e do Caribe em Ciências da Saúde)</i>	Ciências da Saúde, América Latina	Aberto e gratuito
<i>MEDLINE (Medical Literature Analysis and Retrieval System Online) (via Pubmed)</i>	Ciências da Saúde	Aberto e gratuito
<i>PEDro (Physiotherapy Evidence Database)</i>	Fisioterapia	Aberto e gratuito
<i>PsycINFO</i>	Psicologia e ciências sociais	Aberto e gratuito
<i>SciELO (Scientific Electronic Library Online)</i>	Multidisciplinar, Brasil e países colaboradores	Aberto e gratuito
<i>Scopus</i>	Multidisciplinar	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)
<i>Web of Science (Institute for Scientific Information (ISI) Web of Knowledge)</i>	Multidisciplinar	Restrito, mas gratuito via assinatura institucional (Periódicos Capes)

em limites (por exemplo, restringir ao termo, explodir), campos de busca (por exemplo, título, resumo, palavras-chave), permutação do termo (por exemplo, truncar o sufixo, prefixo com uso de símbolos como * ou \$) e uso de operadores booleanos (AND, OR, NOT). Cada base de dados disponibiliza orientações e tutoriais sobre como fazer buscas e utilizar adequadamente os recursos disponíveis em cada base. A escolha de termos limites e combinações afeta diretamente os resultados da estratégia em cada uma das bases escolhidas.

Considerando as especificidades de cada base de dados, recomenda-se iniciar a estratégia de busca por uma base de dados, geralmente a maior ou a mais importante e, se possível a mais familiar, e da forma mais simples possível. Usualmente, inicia-se com os termos mais importantes, como a população ou intervenção/exposição isoladamente e progressivamente avalia-se o resultado de adicionar um termo ou uma combinação de sinônimos de outro componente da pergunta de pesquisa.

Em um primeiro momento, pode-se pensar que o uso das palavras relacionadas a todos os itens do acrônimo PICOS seja a forma mais sensível de busca, entretanto esta não é a melhor estratégia. Especialmente, deve-se evitar o uso do desfecho na busca, pois o desfecho procurado pode não ter sido o desfecho primário (ou principal) do estudo e sim um desfecho secundário (ou adicional) e, por essa razão, não constar no resumo do artigo. A busca pelo desfecho nesse caso deve retornar um estudo que não seja potencialmente útil para a revisão. Além disso, o uso de controle na busca somente deve ser inserido quando se tratar de algo pontual e definido, para não deixar de incluir estudos que tenham utilizado controles diferentes daqueles originalmente pensados na PICOS.

O uso de descritores apropriados para cada base de dados (aqueles que encontram-se na árvore de hierarquia de indexação do descritor) auxiliam na busca. Os *Entry Terms* correspondem aos termos que, se localizados no resumo, geram a marcação deste estudo com o MeSH term relacionado. Os *Entry Terms* são uma excelente forma de recuperar o maior número de sinônimos a serem inseridos na construção da busca. Este processo é parcialmente manual, de forma que não há uma garantia inequívoca de que qualquer resumo sobre um determinado

assunto utilizará necessariamente o descritor correspondente. Uma dica importante nesta etapa é salvar a busca criada na base de dados em um editor de texto para que futuramente seja possível replicar a busca exata e atualizar a lista de estudos.

Tendo abordado estes aspectos, utilizar os *Entry Terms* seria suficiente para localizar todos os artigos de interesse? A resposta é não. Nem sempre os trabalhos utilizam os descritores indexados. Por isso, a importância de algum membro da equipe com expertise na área, pois ele poderá ajudar a definir outros sinônimos de busca.

Em revisão sistemática usualmente não se utiliza restrição de tempo, bem como se recomenda que a busca não seja restrita a um idioma. Além disso, a tradução dos termos para o inglês americano e britânico é aconselhável, por exemplo, “fogachos” é traduzido para o inglês britânico como “hot flushes”, enquanto que para o inglês americano é “hot flashes”. O filtro de tipo de estudo pode ser utilizado ao final, mas é recomendado também incluí-lo na estratégia de busca.

Depois de testada, amadurecida e aprovada para a primeira base de dados, a busca pode ser progressivamente “traduzida” para outras bases de dados. O revisor que busca um conjunto abrangente de estudos primários deve abordar as outras bases de dados e, às vezes, muitas outras. Além disso, uma mesma citação pode estar indexada em mais de uma base de dados, ou seja, duplicada. Todo o processo de busca deve ser registrado e reportado no relato final da revisão sistemática.

O guia PRESS (*Peer Review of Electronic Search Strategies*) (www.cadth.ca/resources/finding-evidence/press) é uma ferramenta que deve ser considerada para desenvolver a estratégia principal de busca. Ele fornece um conjunto de recomendações sobre as informações que devem ser usadas por bibliotecários e outros especialistas em informação quando são solicitados a avaliar essas estratégias de busca eletrônica.

Conhecer a literatura que se deseja revisar também é importante. Alguns artigos podem ser utilizados como sentinelas da estratégia, verificando se esse estudo potencialmente elegível está sendo recuperado na busca a cada etapa de construção da estratégia de busca. Apesar de encher os olhos, uma estratégia grande e cheia de combinações não significa maior qualidade ou sensibilidade da estratégia. Na verdade, é mais fácil se perder ou negligenciar erros com uma estratégia muito complexa. É preciso também alcançar equilíbrio entre sensibilidade e especificidade da estratégia, visando uma opção que represente um número razoável de estudos para a próxima etapa de seleção, recuperando estudos potencialmente elegíveis. Isso implica em uma atividade que depende da experiência e julgamento do revisor. Executar as atividades minuciosamente faz toda diferença.

Literatura cinzenta

A literatura cinzenta consiste em dados não publicados formalmente em fontes como livros, revistas e periódicos eletrônicos, em que a busca pode ser realizada de forma manual ou eletrônica. Os documentos denominados como literatura cinzenta incluem resumos de congressos, relatórios técnicos de pesquisa, preprints, e-prints, trabalhos de conclusão de curso, dissertações, teses, publicações governamentais (emendas, projetos, leis, manuais), anais de eventos científicos, estudos não registrados, registro de ensaios clínicos (revisões de estudos de intervenção precisam considerar essa fonte), lista de referências, palestras, conjuntos de dados, mídia audiovisual, patentes, blogs, fóruns, especialistas na área (contato pessoal com especialistas no tema para identificação de outros estudos e fontes potencialmente elegíveis).

Algumas iniciativas buscam compilar dados de literatura cinzenta para facilitar o processo de busca e acesso, dentre elas destacam-se o Open Grey (www.opengrey.eu), GreyNet International (www.greynet.org), Grey Literature Report (www.greylit.org), além do Registro Brasileiro de Ensaio Clínicos (www.ensaiosclinicos.gov.br). Também pode-se utilizar ferramentas como Google Scholar (scholar.google.com.br/), TRIP (Turning Research into Practice) (www.tripdatabase.com/), ProQuest Dissertations & Theses (about.proquest.com/) e Coleta Capes

que possui um catálogo de teses e dissertações (sucupira.capes.gov.br/sucupira). Também há um manual da Agência Canadense de Avaliação de Tecnologias em Saúde (www.cadth.ca/resources/finding-evidence/grey-matters), que traz orientações de busca em fontes de literatura cinzenta.

Atualização da busca

Nas fontes eletrônicas em geral é possível configurar atualizações automáticas para serem enviadas por e-mail. A equipe define então a logística para sistematicamente avaliar esses resultados após a conclusão da seleção principal. Podem ser estabelecidas algumas alternativas, como a triagem das atualizações automáticas por um dos revisores ou o julgamento de estudos potencialmente elegíveis pelos membros responsáveis pela seleção. Ao final, roda-se novamente toda a busca e realiza-se a seleção pareada.

À medida que são divulgados novos estudos, recomenda-se importar os resultados para a plataforma de seleção e periodicamente selecionar os estudos e incorporá-los na revisão em andamento. Lembre-se: a revisão é um processo sistemático.

Etapa 3: Seleção dos estudos

O que foi abordado até aqui é apenas o início do trabalho de um revisor. Como a estratégia de busca é elaborada para ser altamente sensível, ela retorna um número muito maior de trabalhos do que apenas os elegíveis para o estudo. Por isso, a próxima etapa envolve a seleção dos estudos, que inclui a importação destes para a plataforma de seleção (que pode ser uma planilha eletrônica ou manual), remoção das duplicações, triagem pelos títulos e resumos e leitura do texto completo dos estudos selecionados para confirmação da inclusão. Preconiza-se que essas etapas sejam realizadas simultaneamente de modo independente por pelo menos dois revisores.

Gerenciadores de referências

Após concluir as buscas, os resultados são exportados em arquivos de acordo com cada base de dados. Geralmente no formato eletrônico (XML, *Extensible Markup Language*), textual (txt,

Text File) ou de gerenciador de referências (RIS, *Research Information Systems*). Idealmente esses arquivos são importados para um software gerenciador de referências (Quadro 4) ou específico para revisão sistemática (Quadro 5), que permitem alguma automatização das atividades, como a identificação de artigos duplicados (quando um mesmo artigo está indexado em duas bases de dados diferentes), identificação de concordâncias e discordâncias entre os avaliadores na seleção. Como alternativa, podem ser usadas planilhas manuais com os arquivos de busca no formato de texto.

Quadro 4. Gerenciadores de referências bibliográficas.

Gerenciador	Característica	Disponibilidade
EndNote	Principais estilos de formatação e filtros de importação disponíveis	Pago
EndNote Web	Opção web e gratuita do EndNote	Gratuito
Mendeley	Disponível em versão web e desktop	Gratuito
Zotero	Integrado com navegadores de internet	Gratuito

Quadro 5. Gerenciadores específicos para revisão sistemática.

Gerenciador	Etapas que podem ser executadas no programa	Disponibilidade
Covidence	Triagem, seleção do texto completo, extração dos dados, avaliação crítica	Pago
DistillerSR	Triagem, seleção do texto completo, extração dos dados	Pago
Rayyan	Triagem	Gratuito
RevMan	Elaboração de revisões da Cochrane	-
RevMan Web	Elaboração de revisões da Cochrane	-
SUMARI	Elaboração de revisões do Instituto Joanna Briggs	-

Dispor tanto de programas gerenciadores de referências quanto específicos de revisão sistemática auxilia na condução e organização das etapas e dos artigos da revisão. Independentemente do gerenciador de referência utilizado, o uso dessa ferramenta é importante durante toda a condução da revisão até sua publicação, garantindo a acurácia das citações no relato

final da revisão sistemática. Toda a equipe deve estar familiarizada com essa ferramenta. Na escolha dos softwares deve ser ponderada a conveniência do uso pela equipe e facilidade de uso.

Triagem dos estudos

A seleção dos estudos envolve duas fases. Na fase I, é realizada a leitura dos títulos e dos resumos por dois revisores independentes, sendo o artigo mantido se selecionado por pelo menos um dos revisores. Os revisores avaliam se o estudo atende aos critérios de elegibilidade previamente estabelecidos no protocolo da revisão sistemática. Mais pesquisadores podem participar desta etapa, mas, neste caso, é importante que cada estudo tenha pelo menos dois votos frente à elegibilidade. Em caso de discordância entre os revisores, por ser uma etapa de triagem, um parecer favorável é suficiente para que o artigo permaneça para a próxima etapa. É importante que a seleção ocorra de forma independente para que as impressões e opiniões dos revisores tenham menor influência de um julgamento prévio. Como há subjetividade, é de se esperar que ocorram discordâncias entre os avaliadores. Pode-se realizar reuniões para calibragem, discussão e consenso ou avaliação por um terceiro avaliador, mais experiente, que opte pela inclusão ou exclusão do estudo.

Na fase II, ocorre a leitura dos textos completos dos estudos selecionados na fase anterior por dois revisores independentes que devem concordar em incluir ou excluir o estudo. Se houver discordância, os revisores devem entrar em consenso ou consultar um terceiro revisor mais experiente. Nesta fase há a confirmação da elegibilidade do estudo. Se a triagem anterior foi pouco específica, esta atividade resultará em muitos estudos para obter o arquivo completo de leitura, o que pode gerar trabalho excessivo ou desnecessário. Alguns grupos também realizam calibrações com os responsáveis pela seleção para melhorar a performance desta etapa. Os motivos de exclusão dos estudos avaliados em texto completo devem ser registrados e explicitados na redação final da revisão sistemática.

A obtenção dos textos completos pode representar custos relevantes para a revisão, pois nem todos estão completa ou gratuitamente disponíveis, o que pode consumir tempo da

equipe para aquisição. O texto completo dos estudos geralmente é obtido no formato PDF (*Portable Document Format*) e podem ser empregadas formas automatizadas de obtenção com auxílio de gerenciadores de referência com essa funcionalidade ou pode-se obter um a um. Recomenda-se verificar os recursos disponibilizados para acesso a periódicos, além de obter orientações de especialistas em informação, como bibliotecários, e consultar os tutoriais do gerenciador. Vários gerenciadores de referências permitem localizar e anexar automaticamente os arquivos PDF após uma varredura na rede em que está conectado, além de integrar com gerenciadores de revisão sistemática.

Alternativas para se obter os artigos completos incluem os repositórios alimentados pelos autores das pesquisas como o ResearchGate (www.researchgate.net); ferramentas que identificam textos abertos como Unpaywall (unpaywall.org/); solicitação direta ao autor do artigo facilitada pelo uso de redes como ORCID (orcid.org), Publons (publons.com/researcher) ou LinkedIn (www.linkedin.com) e; pela iniciativa de acesso Sci-Hub (sci-hub.tw), criado pela ativista Alexandra Elbakyan, que foi considerada uma das 10 personalidades científicas de 2016 pela revista Nature.

O mais importante na seleção é garantir que os estudos que cumprem os critérios de elegibilidade sejam realmente identificados e incluídos. A busca e a seleção pelos avaliadores devem ser bem feitas para garantir a inclusão dos estudos que de fato devem ser incluídos. Confirmar se os avaliadores tiveram uma concordância adequada, avaliada pelo índice de concordância Kappa por exemplo, é uma alternativa de análise secundária, pois não estará representando uma boa seleção de estudos, uma vez que ambos revisores podem concordar em um julgamento errado.

Toda a estratégia de seleção dos estudos deve ser apropriadamente registrada: quantidade de artigos recuperados na busca, duplicações removidas, estudos selecionados, incluídos, excluídos e seus motivos. Esse registro será um dos resultados que a revisão sistemática deverá apresentar e, por esse motivo, toda acurácia deve ser garantida. O adequado registro e detalhamento dos procedimentos e resultados contribui para transparência e reprodutibilidade da revisão. Além de todas as questões

metodológicas e de recursos básicos, é necessário observar se os artigos obtidos fornecem todos os dados necessários para elaboração da revisão sistemática.

Por isso, antes de se enveredar na revisão sistemática, deve-se ter certeza de que há artigos científicos sobre o tema. Para que uma revisão seja viável é necessária literatura elegível disponível. Só é possível revisar o que já foi pesquisado e publicado sobre o assunto e, idealmente, que tenha semelhança na pergunta PICOS. Bem como a questão deve representar um dilema científico e verificar se já existe alguma revisão sistemática sobre o tema que se planeja estudar. Se a resposta já está disponível e tem boa qualidade não é necessário novo esforço. Na disponibilidade de revisões sistemáticas, ponderar se a nova revisão de fato trará contribuição relevante, seja pela disponibilidade de novos estudos com tamanho de amostra considerável ou pela aplicação de técnicas (metanálise) que permitirão novos achados e conclusões. Ter uma proximidade com a literatura a ser revisada, bem como confirmar que a dúvida na área existe e necessita de uma resposta, são pressupostos esperados para embasar a necessidade da condução de uma revisão. Essa proximidade permitirá identificar, por exemplo, se os estudos individuais têm resultados imprecisos ou divergentes entre si e melhor embasar a necessidade da pesquisa.

A motivação da equipe para fazer a revisão é extremamente importante. Ter clareza sobre a motivação em se fazer uma revisão sistemática contribui para a avaliação sobre a necessidade dessa pesquisa. Se a motivação da equipe de revisores se afasta de uma investigação acadêmica e científica que precisa ser respondida, há grandes chances dessa revisão ser redundante ou conflituosa. Argumentos frágeis para iniciar uma revisão sistemática e que possivelmente resultarão em uma contribuição contestável ou falha dessa revisão sistemática, tanto no âmbito individual como institucional, incluem: aumentar o número de publicações ou citações, obtenção de prestígio, substituição de um experimento que não deu certo, abandono de uma pesquisa com resultado não desejável, promoção de uma técnica ou produto com interesse comercial, disseminação de incertezas sobre efeitos danosos de produtos ou técnicas, fragilização de medidas regulatórias e de saúde pública e, aumento do número de publicações favoráveis.

Por fim, é imprescindível que a equipe conheça o assunto e os métodos. É fundamental que a equipe domine as expertises mínimas do método de pesquisa e conheça o assunto em análise. O planejamento adequado é essencial. Lembre-se de que há guias disponíveis, o que é uma “mão na roda” para o desenvolvimento da revisão sistemática. No entanto, é importante a disponibilidade de capital intelectual na equipe.

Etapa 4: Extração dos dados

Na extração coleta-se, de forma padronizada, todas as informações relevantes dos estudos, tanto para sua caracterização quanto para a avaliação dos desfechos de interesse. Essa etapa deve ser executada por duas pessoas de forma independente ou por conferência de pesquisador experiente após extração realizada por um outro investigador. Planilhas ou formulários de extração devem ser previamente planejadas e testadas.

Independentemente do processo, formulário e técnicas adotadas na fase de extração de dados, é de suma importância manter a transparência sobre as escolhas metodológicas e os impactos nos dados utilizados na metanálise.

Estudos versus publicações

Nas revisões sistemáticas, os estudos são o objeto de pesquisa e são identificados a partir da recuperação de seus registros ou publicações de seus resultados. Portanto, é importante diferenciarmos o estudo da publicação, os quais podem inclusive não refletir o mesmo número, uma vez que um mesmo estudo pode originar mais de um registro ou publicação. Antes de iniciar a extração dos dados, é importante identificar e vincular os artigos que se referem ao mesmo estudo. Gerenciadores de metanálise permitem vincular publicações pertencentes a um mesmo estudo, o que facilita a etapa seguinte, de extração de dados.

Definindo as variáveis

As variáveis que serão extraídas dos estudos e incluídas na revisão sistemática devem ser definidas previamente e, para isso, deve-se ter em mente quais informações dos estudos

selecionados serão úteis na revisão, para então extraí-los em um formulário específico, como por exemplo, uma planilha de Excel. O procedimento evita retornos desnecessários e frequentes ao texto original do artigo e facilita a condução das análises por consolidar todos os dados em um único local ou arquivo.

Em geral, coletam-se dados sobre autor, ano de publicação, tipo de estudo, participantes (número, média de idade, proporção de cada sexo), detalhes da exposição ou intervenção, desfecho (e como ele foi classificado), tempo de seguimento e perdas. É importante coletar também características de base dos participantes que podem ter influência no desfecho, para que seja possível avaliar a chamada heterogeneidade clínica (variação das características de base dos participantes entre os estudos). É recomendado apresentar na seção de resultados do artigo, uma tabela com as características individuais dos estudos incluídos.

Apesar de não ser um item essencial como no processo de seleção, é desejável que a coleta de dados também seja realizada por dois ou mais pesquisadores independentes. Outra sugestão, na baixa viabilidade de uma extração pareada e independente, é de que um pesquisador realize a coleta dos dados e outro faça a conferência desses dados. Ressalta-se que essa última situação é mais aplicável a grupos experientes no processo.

Ao extrair informações sobre os desfechos, é importante considerar a natureza das variáveis, se numérica ou categórica. Para desfechos contínuos, como glicemia de jejum medida em mg/dL, ou tempo de internação, medido em dias, devem ser coletados o número de amostra, a média e o desvio-padrão de cada grupo de participantes. Em desfechos categóricos, como frequência de diabéticos, os dados coletados são usualmente o número de pessoas que tiveram o desfecho (valores absoluto (n) e relativo (n%)) e o total de sujeitos em cada grupo.

Dados ausentes: e agora?

É comum, infelizmente, publicações e registros não apresentarem os dados de interesse para a revisão sistemática. Nestes casos, é necessário recorrer ao contato com o autor ou a transformações dos números para viabilizar a síntese dos dados.

Quando for contemplada uma síntese quantitativa dos resultados, pode ser necessário escrever aos autores dos estudos primários e solicitar dados que não tenham sido originalmente incluídos na versão publicada. Também, em algumas situações por exemplo, há somente o erro padrão de uma medida contínua, sendo necessário calcular o desvio padrão para que a metanálise seja possível. Mas, há ocasiões em que as tentativas de contato com autores são frustradas. Para evitar excluir o estudo da metanálise devido aos dados ausentes, pode-se fazer uso da imputação desse dado a partir de técnicas estatísticas.

Através da imputação dos dados faltantes são criados conjuntos de dados artificialmente completos que podem ser analisados por técnicas estatísticas tradicionais. Para obtenção destes valores plausíveis, cada dado ausente é imputado "m" vezes, com reposição, gerando bancos de dados completos. Os bancos obtidos são analisados separadamente e os "m" resultados obtidos são combinados na análise final. **Importante!** Através da imputação de dados pode-se subestimar ou superestimar a variabilidade da variável imputada, gerando intervalos de confiança mais estreitos do que o esperado e impossibilitando de levar em consideração a variabilidade que possa existir entre diferentes imputações. Também existem técnicas de imputação múltipla, que utilizam a probabilidade, disponíveis em *softwares* e pacotes estatísticos, fazendo uso de regressões, simulações, inferências bayesianas e até algoritmos de aprendizado de máquina (*machine learning*) para a predição dos valores a serem imputados.

Não há consenso no limite do número de dados que podem ser imputados, mas recomenda-se até 10% da amostra completa. É muito importante prudência na adoção da imputação. Essa estratégia deve ser adotada apenas na impossibilidade de obtenção dos dados de outra forma. Do mesmo modo, sugere-se fortemente que a adoção ou não dos valores imputados seja motivo de análises de sensibilidade de uma metanálise e que seja descrito de maneira transparente.

Registro da extração de dados

Da mesma forma que a seleção dos estudos, é boa prática o registro seguro, organizado e compartilhado dos dados extraídos. Para essa tarefa, usualmente se trabalha com planilhas,

formulários eletrônicos ou softwares específicos, como o Microsoft Excel e o Google forms. Criar um dicionário de variáveis e detalhar procedimentos e transformações adotadas auxilia nesse processo. Também é indicado a criação de formulários de fácil manipulação e compartilhamento, mas de forma segura. A informação coletada deve representar fielmente o que foi relatado nos estudos incluídos de maneira estruturada e organizada. Um bom planejamento dessa etapa permitirá identificar os dados necessários para a síntese quantitativa.

Ciência aberta

A ciência aberta é um conceito “guarda-chuva” (Figura 5) que compreende diferentes níveis de abertura, propriedade e disponibilidade de dados para tornar as publicações transparentes, robustas e reproduzíveis. Relacionado às revisões sistemáticas, destaca-se a revisão por pares aberta (*open peer review*), onde os pareceristas e autores são identificados no processo de revisão por pares para publicação; a publicação de artigos com acesso aberto (*open access*) e; os dados abertos, quando ocorre a disponibilização gratuita dos dados criados.

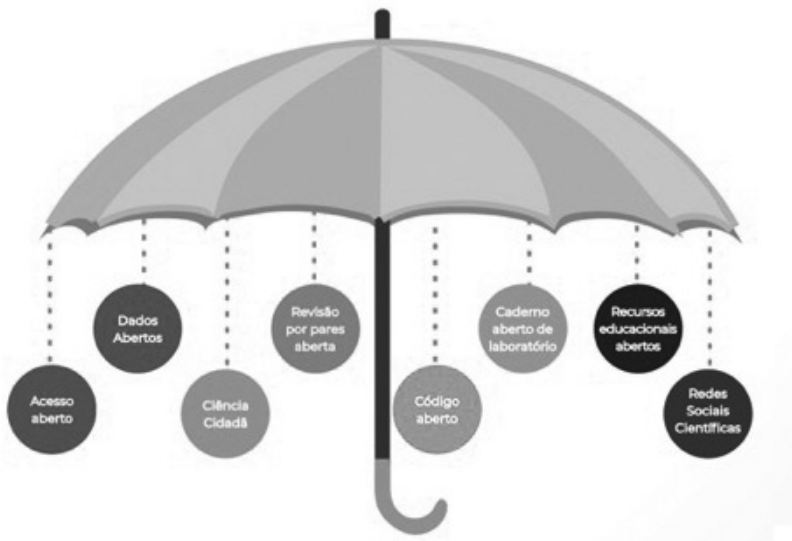


Figura 5. Ciência aberta. Fonte: Fiocruz.

A reprodutibilidade da pesquisa científica por meio do compartilhamento de bancos de dados utilizados nas metanálises é cada vez mais desejável. Tal prática, além de tornar o trabalho transparente e reprodutível, auxilia a recondução futura de uma metanálise por outros membros ou grupos de pesquisa na hipótese de surgirem novas evidências, facilitando o processo de atualização por diferentes grupos, além de análises adicionais. Em alguns periódicos científicos a disponibilização pública do banco de dados já é uma solicitação rotineira aos autores no momento da submissão do artigo científico.

Para auxiliar o compartilhamento público de bancos de dados, já existem plataformas gratuitas de repositórios, como o Data Mendeley (data.mendeley.com/), que permite a disponibilização do banco de dados de forma segura com a geração de um número de identificação único (D.O.I.) que pode ser referenciado em uma eventual publicação científica, e o Github (github.com/) da Pubmed, já bastante disseminado no meio da disponibilização de códigos, arquivos e bancos de dados de softwares abertos.

Vários procedimentos relacionados à ciência aberta já são adotados na revisão sistemática, como registro do protocolo, disponibilização da estratégia de busca, explicação da exclusão de uma pesquisa em particular. O que ainda pode avançar é no compartilhamento dos dados produzidos na revisão sistemática, incluindo, por exemplo, o banco de dados da extração, detalhes da avaliação crítica de cada artigo incluído, cálculos ou imputações realizadas para dados ausentes, comandos e sintaxes adotadas para análises estatísticas.

Neste sentido, é importante tomar cuidado com as chamadas revistas e conferências predatórias. Revistas consideradas predatórias são aquelas que se dispõem a publicar artigos científicos sem submetê-los a uma acurada revisão por pares, desde que ocorra o pagamento da taxa para publicação. Uma forma de evitá-las é consultando o Capes Qualis (sucupira.capes.gov.br/sucupira/) ou outras ferramentas de avaliação de periódicos.

Etapa 5: Avaliação da qualidade metodológica dos estudos

O passo seguinte da revisão sistemática é a avaliação da qualidade metodológica dos estudos incluídos. Esta avaliação indicará a confiança que se terá nos resultados da revisão. O rigor metodológico é um aspecto que não deve apenas ser considerado na elaboração de uma revisão sistemática, mas também na condução dos projetos de pesquisa e na leitura de qualquer estudo científico.

Ao realizar a leitura de uma revisão sistemática, é importante questionar se: a qualidade metodológica foi analisada e os estudos foram valorizados da maneira adequada? Como assegurar que um estudo menor, mas desenvolvido de forma perfeita, receba o valor que merece em relação a um estudo maior cujos métodos são adequados, mas susceptíveis à erros?

Para auxiliar a responder essas perguntas, existem listas de critérios para analisar se um artigo deve ser rejeitado com base na sua metodologia. Neste sentido, o aspecto chave é o grau em que as falhas metodológicas causam vieses nos resultados da revisão sistemática.

Limitação metodológica

Limitação metodológica é qualquer erro (no desenho, condução, análise ou interpretação dos resultados) que distancie o resultado observado do verdadeiro. As limitações metodológicas reduzem a qualidade do estudo. Algumas destas limitações são comumente identificáveis nos desenhos de estudo. Em ensaios clínicos randomizados, por exemplo, é importante ater-se ao tamanho amostral, alocação aleatória (randomização), mascaramento (cegamento), seguimento completo, sigilo de alocação e análise por intenção de tratar. A análise de intenção de tratar (*intention to treat*, ITT) é o método de análise estatística realizada para os dados coletados durante o estudo, que prevê que todos os doentes alocados a um grupo de tratamento sejam seguidos, avaliados e analisados, independentemente do cumprimento do tratamento inicialmente planejado.

Nos estudos observacionais, os pontos críticos são poder estatístico, direcionalidade temporal (retrospectivo ou prospectivo), perda de seguimento, tempo de seguimento e presença de fatores de confusão. Para estudos de acurácia diagnóstica, observa-se o espectro da doença (estágios de gravidade da doença), uso de padrão de referência (uso de padrão não apropriado), cegamento dos avaliadores quanto ao resultado do padrão de referência e tempo de realização entre o padrão de referência e o teste índice (teste novo).

Qualidade metodológica

A avaliação da qualidade metodológica inclui a avaliação do risco de viés de cada estudo frente a critérios metodológicos previamente estabelecidos, visando identificar a presença de vieses de seleção, aferição e fatores de confusão. Para isso, deve-se identificar quais as ferramentas mais apropriadas de acordo com o delineamento do estudo.

Lembre-se de que revisões sistemáticas são utilizadas para orientar processos de tomada de decisão. Uma forma de distinguir revisões sistemáticas das outras formas de síntese é averiguar se houve algum processo de apreciação crítica dos estudos que foram incluídos. Essa avaliação minuciosa é conhecida como avaliação do risco de viés.

A avaliação da qualidade metodológica dos estudos incluídos apresenta dois aspectos importantes:

- 1) a maior parte das evidências consideradas na síntese são provenientes de artigos publicados em periódicos científicos. Assim, esses estudos já passaram por inúmeras avaliações, seja por comitê de ética em pesquisa, banca de mestrado ou de doutorado e revisão por pares. Independente do prestígio do periódico, esses estudos terão mais uma avaliação;
- 2) a avaliação crítica destes estudos é diretamente dependente da qualificação (expertise) dos revisores. Revisores pouco experientes têm abstrata percepção do impacto que eventuais desvios na condução do estudo têm sobre os resultados.

Os autores de revisões sistemáticas devem julgar como eventuais vieses influenciam os resultados encontrados, ou seja, se estes erros podem subestimar ou superestimar os achados. A forma de investigar o risco de viés depende da pergunta de pesquisa e precisa ser previamente especificada no protocolo da revisão sistemática.

Guias para avaliação da qualidade dos estudos

Há diversos guias disponíveis para avaliar criticamente os estudos incluídos em uma revisão sistemática. Geralmente, estão apresentados como listas de verificação, escalas ou checklists. Não há consenso sobre o melhor método de verificação da qualidade metodológica devido à grande quantidade de escalas e listas disponíveis. No quadro 6, estão apresentados alguns destes guias.

A ferramenta *Risk of Bias* possui seis domínios:

- 1) randomização (alocação aleatória dos participantes da pesquisa para os grupos de estudo, intervenção e controle);
- 2) sigilo de alocação (evita que o pesquisador antecipe para qual grupo um determinado paciente será alocado);
- 3) cegamento (o investigador trata ou avalia sem saber para qual grupo o paciente será alocado);
- 4) desfechos incompletos (perda de dados, perda de seguimento, exclusões, análises não realizadas pelo princípio da ITT ou qualquer perda diferencial entre os grupos);
- 5) relato seletivo (desfechos e análises não esperadas e/ou planejadas e desfechos que foram considerados no protocolo do ensaio clínico randomizado, mas foram omitidos da publicação) e;
- 6) outras fontes de viés (interrupção prematura do estudo, contaminação (mudança de tratamento), diferenças da linha de base).

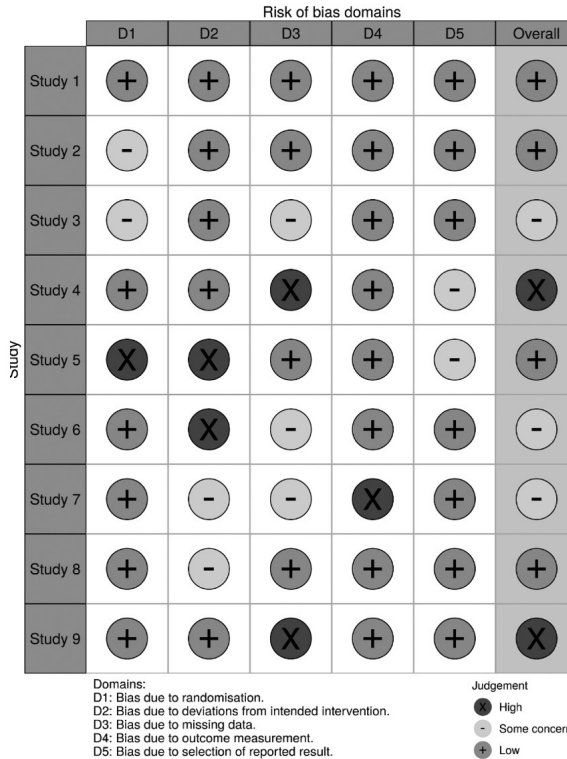
Cada domínio pode ser classificado em três categorias (Figura 6): Baixo risco de viés (existe pouca evidência de que vieses nesse domínio tenham influenciado de modo significativo os resultados do estudo); Alto risco de viés (existe evidência de que falhas nesse domínio podem ter influenciado de modo significativo os resultados

Quadro 6. Guias para avaliação da qualidade metodológica dos estudos.

Tipo de estudo	Guia de avaliação metodológica
Ensaio clínico randomizado	<i>Revised tool for Risk of Bias in randomized trials (RoB 2.0 tool)</i> (sites.google.com/site/riskofbiastool)
Ensaio clínico não randomizado	<i>Risk Of Bias in Non-randomized Studies of Interventions (ROBINS-I tool)</i> (sites.google.com/site/riskofbiastool)
Estudo de coorte	<i>Newcastle-Ottawa Scale</i> (www.ohri.ca/programs/clinical_epidemiology/oxford.asp)
Estudo caso-controlado	<i>Newcastle-Ottawa Scale</i> (www.ohri.ca/programs/clinical_epidemiology/oxford.asp)
Estudo transversal	<i>Joanna Briggs critical appraisal tools</i> (joannabriggs.org/critical-appraisal-tools)
Estudo de acurácia diagnóstica	<i>Revised toll for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)</i> (DOI: 10.7326/0003-4819-155-8-201110180-00009)
Estudos pré-clínicos	<i>Systematic Review Centre for Laboratory animal Experimentation (SYRCLE)</i> (DOI: 10.1186/1471-2288-14-43) e <i>Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES)</i> (www.dcn.ed.ac.uk/camarades/)
Estudos de avaliação econômica em saúde	<i>Drummond 1996</i> (DOI: 10.1136/bmj.313.7052.275) e <i>Nunes da Silva et al., 2014</i> (DOI: 10.5123/S1679-49742016000100023)
Diretrizes clínicas	<i>Appraisal of Guidelines for Research and Evaluation (AGREE)</i> (www.agreetrust.org/)
Revisões sistemáticas	<i>A Measurement Tool to Assess Systematic Reviews (AMSTAR 2)</i> (amstar.ca/Amstar_Checklist.php), <i>Risk Of Bias In Systematic Reviews (ROBIS)</i> (www.bristol.ac.uk/population-health-sciences/projects/robis/) e <i>visualization tool for risk of bias assessments in a systematic review (robvis)</i> (sites.google.com/site/riskofbiastool)
Relatórios de avaliação de tecnologias em saúde	<i>International Network of Agencies for Health Technology Assessment (INAHTA)</i> (www.inahta.org/hta-tools-resources/briefs)

do estudo, o que reduz a credibilidade dos resultados) e; Risco incerto (os detalhes disponíveis não são suficientes para avaliar os aspectos metodológicos).

A



B

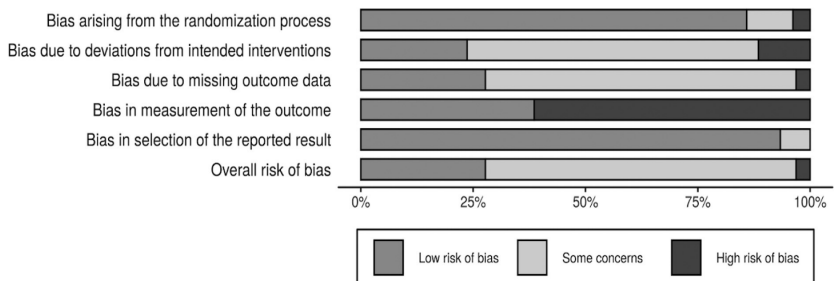


Figura 6. Gráficos da qualidade metodológica avaliada pelas ferramentas RoB 2.0 e ROBINS-I. (A) Gráfico de “semáforo” da avaliação dos domínios para cada estudo e (B) gráfico de barras da avaliação global dos estudos para cada domínio. Fonte: *Risk of Bias*.

Dentre as vantagens, destaca-se a possibilidade de se avaliar os domínios relacionados ao rigor metodológico com a possibilidade de utilizar informação externa para os julgamentos, dando ênfase para os aspectos de qualidade metodológica e não para a qualidade da escrita; transparência da avaliação e; possibilidade de avaliação do risco de viés em desfechos específicos (subdomínios). Porém, limita-se pela subjetividade, sendo necessário consenso entre dois investigadores independentes ou a opinião de um terceiro avaliador; tempo necessário para avaliação e; as inferências obtidas não são válidas para a qualidade geral da evidência e força da recomendação, que será avaliada pelo *Grading of Recommendations Assessment, Development, and Evaluation* (GRADE).

O QUADAS-2, que avalia o rigor metodológico de estudos de acurácia, considera quatro domínios:

- 1) seleção dos participantes (viés de seleção);
- 2) teste índice utilizado (novo método);
- 3) padrão de referência utilizado (acurácia do padrão-ouro) e;
- 4) fluxo e tempo (descrever os pacientes que não foram avaliadas pelo teste índice e/ou padrão de referência ou quem foi excluído das análises, além de descrever o intervalo de tempo entre as aferições).

Cada domínio é avaliado em termos de risco de viés, exceto o domínio fluxo e tempo que é avaliado em termos de aplicabilidade. A classificação ocorre em três categorias: ☺ Baixo risco de viés, ☹ Alto risco de viés e (?) Risco incerto. As vantagens são a padronização, transparência e flexibilidade da ferramenta. Como limitações, subjetividade, principalmente relacionada a aplicabilidade, onde não há questões norteadoras que guiem o julgamento.

A AMSTAR avalia onze aspectos de uma revisão sistemática:

- 1) presença de um projeto de pesquisa (protocolo);
- 2) seleção de estudos e extração de dados por dois revisores independentes;
- 3) busca completa da literatura;

- 4) emprego do "status" da publicação como critério de inclusão (publicado ou não (literatura cinzenta) e idioma);
- 5) apresentação detalhada dos estudos incluídos e excluídos;
- 6) fornecimento das características dos estudos incluídos;
- 7) avaliação da qualidade metodológica dos estudos incluídos;
- 8) uso apropriado da qualidade científica na elaboração das conclusões;
- 9) uso apropriado de métodos para combinar os achados dos estudos originais (metanálise);
- 10) avaliação da probabilidade de viés de publicação e;
- 11) documentação da existência de conflitos de interesse.

Para cada item pode-se obter quatro respostas: sim (1 ponto), não (0 ponto), não se aplica (0 ponto) ou não é possível responder (0 ponto). Soma-se os pontos e, de um total de 11 pontos, classifica-se a qualidade metodológica da revisão sistemática em baixa (0-4 pontos), moderada (5-8 pontos) ou alta (9-11 pontos). Dentre as vantagens, é uma ferramenta simples, rápida e de fácil interpretação, além de apresentar uma boa concordância inter-observador. Limita-se pela não ponderação de cada item e pelo julgamento apenas do que foi relatado e do rigor metodológico.

Qualidade metodológica versus qualidade da evidência

Além da qualidade metodológica dos estudos incluídos em uma revisão sistemática, pode-se e deve-se avaliar a qualidade da evidência dessa revisão. Um estudo secundário bem feito não significa que a qualidade da evidência é alta. Uma revisão sistemática de boa qualidade não corrigirá os erros metodológicos existentes nos estudos primários. A qualidade da revisão sistemática indicará o quanto ela é confiável e adequada para as conclusões apresentadas. A qualidade da evidência avalia a confiança nos achados e parte do pressuposto que a síntese da evidência foi adequada. Ao avaliar o risco de viés de estudos secundários verifica-se os processos de coleta das informações, análise e interpretação e não se a evidência é de boa qualidade.

Etapa 6: síntese dos dados (metanálise)

Qual é a sensibilidade dos resultados em relação à maneira como a revisão sistemática foi conduzida? E se os autores tivessem mudado os critérios de inclusão? E se tivessem excluído trabalhos não publicados? E se a análise da qualidade dos estudos tivesse sido diferente? E se os estudos de qualidade metodológica inferior tivessem sido incluídos ou excluídos? A exploração dos “E se?” é conhecida como análise de sensibilidade. Se mudanças ocorrerem na forma como é realizada a síntese dos dados, é necessário cautela na interpretação das evidências obtidas pela revisão sistemática.

A síntese dos dados é a etapa em que ocorre a sumarização dos resultados, preferencialmente por meio de metanálise (quando possível), com investigação da heterogeneidade, análises de subgrupo, metarregressões e avaliação do viés de publicação. A síntese dos dados objetiva responder à pergunta de pesquisa, seja ela qualitativa, quando não conseguimos aplicar métodos estatísticos, ou quantitativa, nos casos em que a metanálise é possível. Essa etapa é muito dependente da qualidade da extração dos dados e da avaliação crítica dos estudos, ou seja, das etapas que precedem a síntese dos dados.

A síntese qualitativa exige criatividade e clareza para traduzir informação de forma compreensível, sem lançar mão de textos longos que podem obscurecer a comunicação dos resultados e a esperada resposta para a pergunta de pesquisa. Nesses casos, a apresentação dos resultados deve realçar o que é comum nos estudos e, preferencialmente, apresentar de forma comparativa os resultados de cada estudo para os desfechos primários e secundários. Tabelas podem ser ferramentas úteis para resumir os resultados de cada estudo, minimizando a necessidade de explicar cada estudo e seu resultado individualmente.

Definição de metanálise

O primeiro aspecto importante é que revisão sistemática e metanálise não são sinônimos. A revisão sistemática garante a transparência e a maior isenção possível do viés de seleção dos estudos, já a metanálise trata de como serão calculados e combinados

os resultados quantitativos (proporções, riscos relativos, diferenças de médias, etc.) dos estudos incluídos na revisão.

A metanálise é o tratamento estatístico do resultado da revisão sistemática, fornecendo uma média ponderada das estimativas observadas nos estudos obtida através do somatório dos valores multiplicado por seus pesos relativos. A execução de uma metanálise requer pelo menos dois estudos primários ou mais. Desta forma, a metanálise é a principal forma de síntese quantitativa dos resultados, específica para cada desfecho.

A metanálise aumenta o poder estatístico, por reunir dados de estudos diferentes que avaliaram a mesma intervenção para condições específicas. Por meio da técnica estatística, os resultados para um mesmo desfecho são combinados e um novo dado é gerado fornecendo a estimativa global ponderada pelos estudos. Esse cálculo é feito com o auxílio de softwares estatísticos, específicos ou não para metanálise. A estimativa de efeito obtida da metanálise é mais precisa, com menor probabilidade de erro aleatório que os estudos primários, por considerar a soma da amostra dos estudos incluídos.

A coleta dos dados é feita pelos pesquisadores, então mesmo uma metanálise apresentada em uma revisão sistemática pode trazer estimativas imprecisas ou incompletas. As metanálises tendem a ser apresentadas de forma padronizada, facilitando a sua compreensão e leitura. Isso se deve, em parte, ao fato dos programas estatísticos utilizados para fazer os cálculos serem padronizados.

Na metanálise, as medidas de frequência ou de associação são genericamente denominadas medidas de efeito ou de estimativa global, e são utilizadas para sumarizar estatisticamente os resultados. Em metanálises de frequência, geralmente, se apresenta como medida de efeito a própria medida de frequência (prevalência ou incidência). A medida de associação nesses casos seria a razão de prevalência e o risco relativo, respectivamente, e, se sumarizados, ter-se-ia uma metanálise de fatores associados ou de risco.

Se a medida de frequência foi o risco, utiliza-se como medida de associação o risco relativo (RR), sendo necessários o número de participantes com o desfecho e o total de cada grupo

para a metanálise. Se a medida de frequência foi a chance ou probabilidade (*odds*), utiliza-se como medida de associação a razão de chance (*odds ratio*, OR), sendo necessários o número total de participantes com e sem o desfecho em cada grupo. Para o hazard ratio (HR), seria necessário o logaritmo do HR e o erro padrão ou intervalo de confiança. Se a medida for contínua, como a média, a medida de associação seria a diferença de médias, sendo necessário o número de participantes, média e desvio padrão do desfecho em cada grupo.

Além de adaptar a informação em números, a metanálise inclui a tabulação das informações relevantes sobre critérios de inclusão, o tamanho da amostra, características basais dos pacientes, taxas de perda/abandono do estudo, resultados dos desfechos primário e secundário de todos os estudos incluídos, o que facilita o processo de síntese da evidência.

Apesar de sua importância, nem toda revisão sistemática fornece dados apropriados para o desenvolvimento de uma metanálise, e sua realização nestes casos pode levar a resultados errôneos. A metanálise permite a avaliação do resultado em relação a diferentes subgrupos, e permite testar hipóteses diferentes das testadas nos estudos primários. Ressalva-se que o resultado da metanálise está sujeito aos vieses dos estudos primários.

Como ler um gráfico de floresta (forest plot)?

O *forest plot* ou gráfico de floresta é a representação gráfica dos resultados de uma metanálise. O *forest plot* apresenta graficamente os resultados de cada um dos estudos primários juntamente com o resultado combinado, mostrando de forma sintética as estimativas e os intervalos de confiança para cada estudo, bem como para a estimativa combinada ou global. Cada estudo é representado e pode ser identificado pelo nome do autor.

É de extrema importância entender como proceder à leitura e interpretação de um gráfico de floresta. No exemplo, uma revisão sistemática desenvolvida por Rodrigues et al. (2008), sobre a eficácia da intervenção familiar para portadores de esquizofrenia em tratamento ambulatorial. O gráfico de floresta

sumariza os resultados das diversas abordagens, tendo recaída como desfecho (definida como recorrência dos sintomas ou aumento da intensidade dos mesmos).

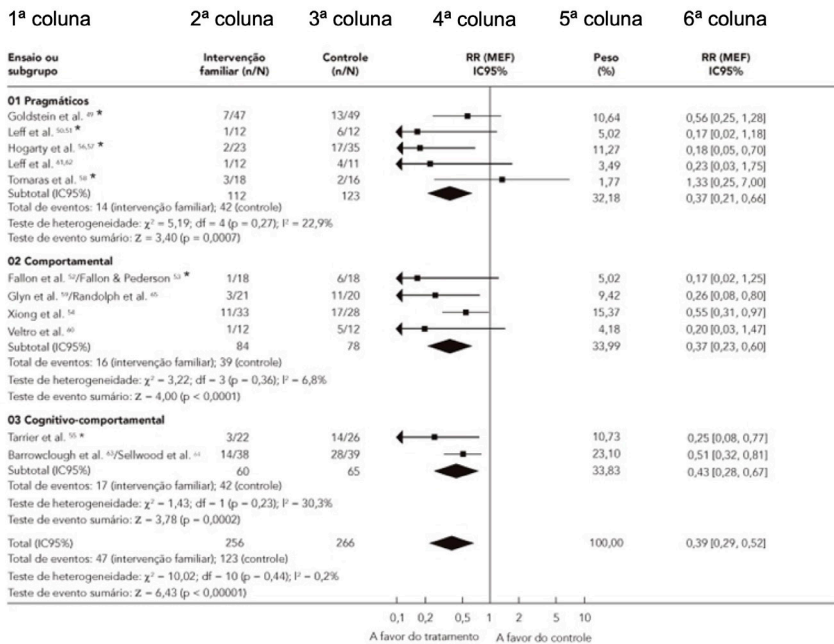


Figura 7. Exemplo de *forest plot*. Fonte: Rodrigues et al., 2008.

Na primeira coluna, temos a identificação dos estudos (geralmente representados pelo autor, ano e pela referência correspondente), listados sequencialmente, um abaixo do outro. Neste caso, subdivididos de acordo com a abordagem utilizada nos ensaios clínicos: métodos pragmáticos, comportamentais ou cognitivo-comportamentais.

Na segunda e na terceira colunas estão representadas as proporções de eventos de interesse (recaída) em cada estudo para os grupos intervenção e controle, respectivamente, temos o número de eventos e o tamanho dos grupos. O valor apresentado na quinta coluna corresponde ao peso de cada publicação para o resultado agregado da metanálise. A quarta e a sexta colunas correspondem ao tamanho do efeito de cada estudo, representado pelo RR e por

seus respectivos intervalos de confiança de 95% (IC95%), de forma gráfica na quarta coluna e de forma numérica na última. A “bolha” no meio de cada linha é a estimativa de ponto da diferença entre os grupos (intervenção e controle), e a largura da linha representa o IC95% desta estimativa. Neste exemplo, o ponto central (“bolha”) de cada linha horizontal representa o RR de cada estudo. No caso da recaída, se o ponto central estiver à esquerda da linha central do gráfico, indica que o tratamento avaliado reduziu a probabilidade de recorrência ou aumento da intensidade dos sintomas. Se o ponto central estiver à direita da linha central, indica que a intervenção familiar piorou a recaída. O tamanho de cada ponto central (“bolha”) indica o peso relativo de cada estudo no resultado final. Esse peso é baseado no número de participantes e no número de eventos. Estudos maiores têm peso maior.

A linha vertical central em “1” corresponde ao RR e é chamada de linha do efeito nulo. Se o desfecho da metanálise for contínuo (diferença entre médias), teremos uma linha vertical centrada em “0”; se diferenças entre riscos (RR/OR), teremos uma linha vertical centrada em “1”. Lembre-se que RR/OR igual à “1” significa que as chances de ocorrência no grupo intervenção e controle são as mesmas. Se a linha horizontal de qualquer estudo não cruzar a linha do efeito nulo, existe 95% de chance de haver uma diferença “real” entre os grupos.

Após cada conjunto de estudos pertencentes a um mesmo subgrupo, encontram-se os resultados unificados (agregados), apresentando o tamanho do efeito comum aos grupos, sendo representado graficamente por um losango (chamado de “diamante”), onde seu centro corresponde ao tamanho do efeito e suas pontas ao IC95%. Ou seja, o ponto central do diamante representa o RR da metanálise e seu tamanho representa o IC (resultado final). Se o diamante estiver à esquerda (benéfico) ou à direita (malefício) da linha vertical, ele é significativo; se ele tocar ou cruzar a linha vertical, não há diferença significativa na metanálise.

Abaixo são apresentados o total de eventos e os valores dos testes de heterogeneidade (Qui-quadrado), medida de inconsistência e teste de hipóteses. No final do gráfico, são apresentadas estas mesmas medidas, unificando os efeitos das três abordagens frente

ao tratamento controle. O diamante representa os dados agregados de todos os estudos incluídos, com um novo IC muito mais estreito, para esse RR de 0,39 (IC95%: 0,29-0,52). Se o IC não contiver o valor neutro (nulo), ou seja, não tocar nem cruzar a linha vertical, estes resultados são considerados estatisticamente significantes. No exemplo, como o losango não se sobrepõe à linha do efeito nulo, pode-se dizer que existe uma diferença estatisticamente significativa.

O valor do Qui-quadrado é um teste estatístico de homogeneidade do tamanho do efeito entre os estudos, isto é, uma medida de consistência do resultado entre os estudos individuais. O valor de z é um teste estatístico da significância do efeito global, isto é, uma medida matemática equivalente à localização e à largura do diamante no gráfico.

Dados individuais dos participantes

Uma alternativa para realizar metanálise é a utilização de dados individuais dos participantes (*individual participant data*, IPD). Para isso, é necessário obter o banco de dados de cada estudo incluído. Dentre as vantagens deste tipo de metanálise, têm-se uma melhor comparabilidade entre os estudos e pacientes incluídos, bem como evita-se vieses como relato seletivo de desfecho. Por depender da obtenção dos dados completos de pesquisas prévias, essa abordagem não é a mais frequentemente observada nas metanálises pois nem sempre se obtêm os bancos de dados.

Metanálise direta

A metanálise direta é a forma mais clássica e comum de combinação dos resultados dos estudos individuais de intervenção ou exposição, sumarizando a medida de associação (risco relativo ou *hazard ratio* para desfechos dicotômicos e diferença de médias para desfechos contínuos) das comparações diretas entre intervenção/exposição e comparador. Se a metanálise é de frequência, como a prevalência, a medida de efeito calculada é de frequência (incidência, prevalência) e não de associação.

Inverso da variância

Compreende-se a variância como medida do grau de "incerteza" ou de "confiança" na estimativa de efeito pontual

(média) de cada estudo, sendo matematicamente o quadrado do erro padrão. Ou seja, quanto maior o valor da variância, mais incertos estamos do resultado médio do estudo.

Com o inverso da variância ($1/\text{variância}$) de cada estudo, temos o raciocínio contrário: quanto maior seu valor, mais preciso o estudo. Assim, adotando o inverso da variância como fator de ponderação os estudos com maior certeza das estimativas pontuais têm mais peso. Esta é uma estratégia usada para dar mais importância aos estudos com resultados mais precisos (estudos com um tamanho de amostra maior, por exemplo). Este método tem alta aplicabilidade, podendo ser usado para combinar dados contínuos ou categóricos e para modelos de efeitos fixos ou aleatórios.

Efeitos fixos e aleatórios

Ao conduzir e, até mesmo ao avaliar criticamente os resultados de uma metanálise, existem dois conceitos importantes que devem ser considerados: modelos de efeito fixo e de efeito aleatório (Figura 8). O modelo de efeitos fixos estima a medida sumária, ou seja, o resultado da metanálise, assumindo um único efeito real (fixo) na população de estudo. Já no modelo de efeitos aleatórios, não existe um valor único de efeito real na população, mas uma distribuição de valores de efeitos possíveis e que, além do acaso, dependem das características dos estudos incluídos.

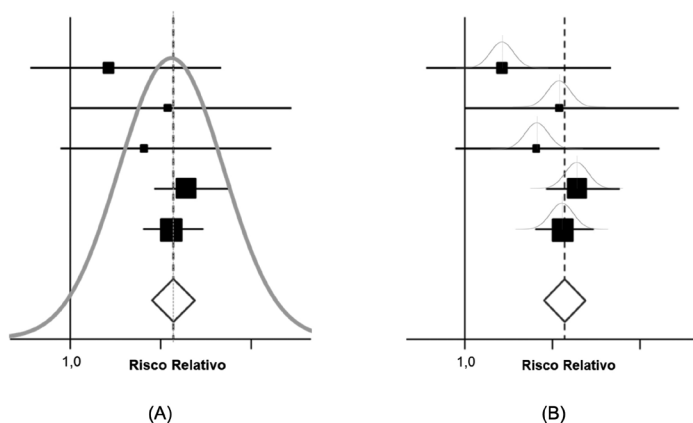


Figura 8. Representação de modelos de efeito fixo (A) e de efeito aleatório (B).

Desta forma, na construção dos ICs das estimativas sumárias, o modelo de efeitos randômicos (aleatórios) considera também a heterogeneidade estatística e terá ICs maiores. Pode-se considerar assim que esse modelo nos informa uma estatística mais conservadora dos efeitos em relação ao modelo de efeitos fixos. Assim, a medida sumarizada possui um IC mais amplo e os estudos menores ganham peso maior do que possuíam em um modelo de efeitos fixos. Isso pode aumentar o impacto do viés de estudos pequenos.

Metanálise em rede

A metanálise em rede (*network meta-analysis*) permite estabelecer uma rede de comparações mais ampla combinando as evidências diretas dos estudos incluídos, o que chamamos de *head-to-head*, com as indiretas por meio dos efeitos relativos de cada tratamento.

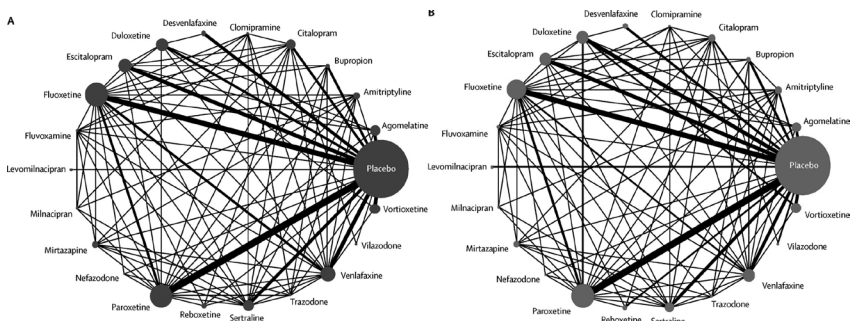


Figura 9. Representação de metanálise em rede da eficácia (A) e aceitabilidade (B) de 21 antidepressivos. Fonte: Cipriani et al., 2018.

A figura 9 exemplifica uma metanálise em rede de comparações de eficácia e aceitabilidade de 21 antidepressivos de Cipriani *et al.* 2018, onde a espessura da linha reflete o número de comparações diretas disponíveis de um conjunto de tratamentos. O tamanho de cada círculo é proporcional ao número de participantes designados aleatoriamente (ou seja, tamanho da amostra). Todas as intervenções dentro dessa rede podem ser comparadas entre si, existindo (direta) ou não (indiretamente) um estudo de comparação para cada uma delas, como a eficácia e a aceitabilidade da

Desvenlafaxina que pode ser comparada indiretamente com a Paroxetina pela Duloxetina. Para que um estudo faça parte da rede é necessário ter pelo menos uma comparação direta (uma linha) com algum dos estudos do conjunto de evidências. Contudo, é importante avaliar se foram garantidos a transitividade (ou similaridade) dos estudos incluídos na comparação.

Outras abordagens de análise

Além do método do inverso da variância, há também os modelos de Mantel Haensel, utilizado para dados dicotômicos e para modelos de efeito fixo e aleatório. É preferencialmente utilizado quando o estudo apresenta poucos eventos e tamanho de amostra pequeno. O método de Peto é utilizado quando a medida de efeito é OR e o modelo de efeito fixo. O método de DerSimonian e Laird é uma abordagem frequentista da metanálise e se baseia em uma média ponderada das medidas de efeito de cada um dos estudos.

Da mesma forma, na metanálise em rede há abordagens variadas. Cada método tem suas aplicações e limitações de acordo com as características dos estudos disponíveis e as variáveis em análise. A escolha de qual método adotar deve ser criteriosa e transparente, podendo inclusive fazer parte de análises de sensibilidade, onde é possível identificar o impacto dessas escolhas sobre as conclusões da metanálise. Deve-se procurar conhecer os métodos disponíveis e aplicá-los em suas análises e interpretações.

Softwares para metanálise

A escolha do *software* para metanálise depende de fatores como a possibilidade de aquisição de licença, nível de domínio de programação e a complexidade dos modelos e seus respectivos gráficos.

Inúmeros softwares estão disponíveis para a realização de metanálise. Alguns softwares de livre acesso e interface gráfica amigável, como o RevMan (training.cochrane.org/online-learning/core-software-cochrane-reviews/revman), disponibilizado pela colaboração Cochrane, permitem a construção dos principais modelos e gráficos da metanálise. Contudo, não permite

a realização de algumas análises menos triviais, como a metarregressão e a metanálise em rede. O programa R (www.r-project.org), um software gratuito e amplamente utilizado para metanálises, requer domínio de programação avançada. O JASP (jasp-stats.org/), uma plataforma amigável para quem tem familiaridade com o SPSS, é um software de acesso livre, bastante flexível, sendo desejável domínio de programação. Por fim, o Stata (www.stata.com), software capaz de realizar a metanálise, apesar de bastante flexível, é pago. Muitos destes softwares possuem tutoriais disponíveis. No entanto, um estatístico pode auxiliar nesta etapa da revisão sistemática.

Heterogeneidade

Um dos principais cuidados para evitar distorções nos resultados da metanálise diz respeito à heterogeneidade entre os estudos. Na metanálise, homogeneidade significa que os resultados de cada estudo individual são compatíveis/comparáveis com os resultados de qualquer um dos outros. Uma forma de evitar resultados heterogêneos é incluir na revisão sistemática apenas os estudos metodologicamente semelhantes, em termos de população, intervenção/exposição, comparador, desfecho e delineamento.

Visualmente, assim como em um histograma para verificar a normalidade de distribuição dos dados, pode-se de forma simples verificar se há heterogeneidade observando se os resultados dos estudos "se conversam", se os seus ICs se sobrepõem parcial ou totalmente. Se não há sobreposição, deduz-se que os resultados divergem estatisticamente e, então, deve-se fazer a análise por outras técnicas.

Da mesma forma, julgar que um conjunto de estudos é heterogêneo com base na sobreposição ou não de seus ICs é subjetivo. Contudo, podem ocorrer variações que são difíceis de serem detectadas ao inspecionar os estudos primários, como mudança de critérios diagnósticos ao longo do tempo e influência do contexto. Para identificar a heterogeneidade nos achados são aplicadas técnicas estatísticas com o propósito de verificar se as diferenças observadas nos resultados são explicadas pelo acaso.

O teste do Qui-quadrado é um dos mais empregados para avaliar a significância da heterogeneidade. Tem sido convencionado um nível de significância mais conservador de $p < 0,100$, em lugar do usual $p < 0,050$. A magnitude da heterogeneidade é averiguada principalmente pelo cálculo do I-quadrado (I^2), que varia de 0 a 100%. Em geral, um I^2 superior a 50% indica heterogeneidade substancial e, acima de 75%, heterogeneidade considerável. Quanto maior a heterogeneidade, maior o questionamento sobre a validade de combinar resultados. Na presença de heterogeneidade, recomenda-se investigar suas causas por meio de análise de subgrupo e metarregressão.

Uma estatística do Qui-quadrado tem, em média, um valor igual aos seus graus de liberdade (número de estudos na metanálise menos um). Um valor de Qui-quadrado muito maior que o número de estudos em uma metanálise nos diz que os estudos que contribuíram para a análise são diferentes uns dos outros em algum aspecto importante. Pode haver, por exemplo, diferenças conhecidas na metodologia (por exemplo, os autores podem ter usado questionários diferentes para avaliar a exposição) ou diferenças clínicas conhecidas nos participantes do estudo (por exemplo, os locais de recrutamento dos participantes apresentavam diferenças quanto a gravidade da doença). Entretanto, pode haver diferenças desconhecidas ou não registradas entre os estudos sobre as quais o investigador pode somente especular até que tenha obtido mais detalhes dos autores do estudo. Outro teste utilizado, mas que apresenta menor poder para detectar se existe heterogeneidade entre os estudos (poucos estudos na metanálise), é o teste Q de Cochran.

Da mesma forma, a análise da heterogeneidade pode ser realizada por um estatístico. Mas explicar essa heterogeneidade (procurar e justificar a heterogeneidade) é um exercício de interpretação e requer bom senso e experiência do investigador. O pesquisador precisa responder de que modo o estudo A diferiu do B e o que os estudos C e D têm em comum para fazer seus resultados se agruparem em um extremo da figura, por exemplo.

Análise de subgrupo

Neste tipo de análise avalia-se os estratos específicos da população recrutada nos estudos incluídos ou em propriedades dos estudos incluídos para identificar sua repercussão nos resultados. Pode-se avaliar, por exemplo, diferentes subgrupos de pacientes (pacientes obesos grau 2 versus pacientes obesos grau 3), de intervenções (métodos pragmáticos, comportamentais ou cognitivo-comportamentais demonstrado na figura 7) ou de estudos (estudos de coorte versus estudos caso-controle). Se houver resultados mais homogêneos, os revisores precisam avaliar a plausibilidade do achado.

Metarregressão

Em estudos primários, os modelos de regressão (análise multivariada) são utilizados para avaliar a relação entre uma ou mais covariáveis com a variável dependente. Da mesma forma, na metanálise, pode-se utilizar a metarregressão, que nos permite avaliar o efeito de múltiplos fatores na heterogeneidade, respeitando-se o mínimo de dez estudos para que a análise tenha validade.

O gráfico da metarregressão recebe o nome de *bubble plot* (Figura 10). No eixo X é apresentada a covariável e no eixo Y a medida de associação de cada estudo. Cada círculo consiste em um estudo e o tamanho do círculo é proporcional ao tamanho do estudo.

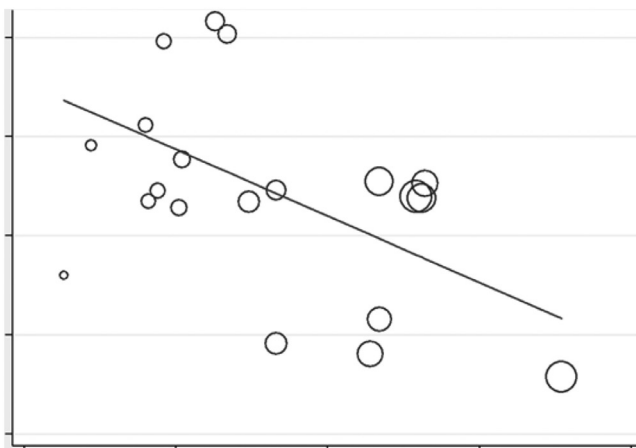


Figura 10. Representação de Bubble Plot.

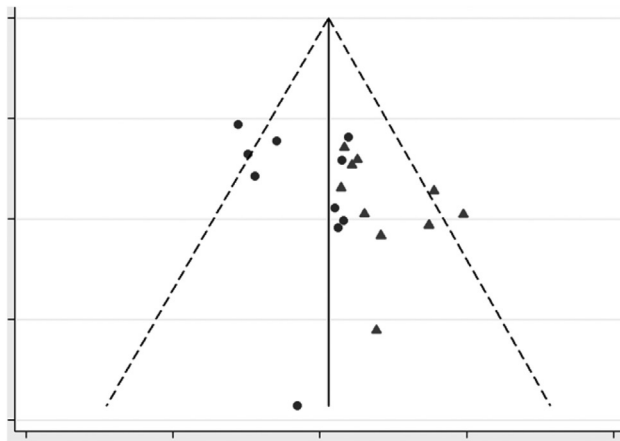
Viés de publicação

O viés de publicação é uma das principais limitações que pode afetar o resultado de uma revisão sistemática. Trata-se da tendência de os resultados publicados estarem sistematicamente diferentes da realidade. Para a avaliação da presença desse viés nas revisões sistemáticas, utiliza-se ferramentas como o gráfico de funil (*funnel plot*) (Figura 11) e testes estatísticos. A utilização dessas técnicas é recomendada para metanálises com dez estudos ou mais e baseia-se em questões de estimativa e de precisão.

Os dados publicados diferenciam-se sistematicamente da realidade quando, por algum motivo, dados são sistematicamente não publicados ou não disponibilizados pela literatura científica. A não publicação de resultados pode ser devida à decisão do autor ou do financiador do estudo, que não submetem para publicação os achados desfavoráveis, ou dos editores de periódicos científicos, que podem não ter interesse em divulgar resultados negativos (sem significância estatística), refletindo uma tendência à distorção da evidência por meio da publicação seletiva.

O gráfico de funil distribui os estudos de acordo com o tamanho amostral e nível de precisão do resultado encontrado. Os estudos pouco precisos, em geral realizados com amostras de tamanho pequeno, poderão encontrar resultados positivos ou negativos (estatisticamente significativos ou não) por influência do acaso. Eles estariam distribuídos simetricamente na parte mais larga do funil. Estudos de maior precisão, em geral em menor número, estariam mais próximos do valor real e situados na parte mais estreita do funil.

No eixo Y do gráfico, temos uma medida de dispersão, o erro padrão, que é influenciado pelo tamanho de amostra do estudo. Quanto maior esse valor, maior é a imprecisão do estudo e mais próximo da base estaria posicionado. No eixo X do gráfico, temos a medida de efeito mensurada na metanálise e a linha central é o resultado da metanálise (estimativa média ou estimativa global). A linha pontilhada representa o IC95%.



Círculo = estudos publicados; Triângulo = estudos não publicados.

Figura 11. Representação de *Funnel Plot*.

Uma vez que estudos imprecisos têm a mesma chance de encontrar resultados positivos e negativos devido à influência do acaso, espera-se que esse funil tenha resultados distribuídos simetricamente no gráfico. O exemplo sugere assimetria, pois os estudos não publicados (triângulos) estão apenas do lado direito, superestimando o valor da linha vertical central.

O uso do gráfico de funil tem algumas limitações, como seu baixo poder de detecção quando existem poucos estudos incluídos na metanálise. Por isso, recomenda-se que ele somente seja utilizado em revisões com mais de dez estudos. Além da interferência de estimativas de efeito relacionadas com o erro padrão ou da presença de heterogeneidade, que produzem falsa assimetria no gráfico.

Os testes estatísticos mais comuns para avaliar viés de publicação são o teste de Egger e o teste de Begg. O teste de Egger é o mais frequentemente utilizado, contudo pressupõe uma distribuição normal dos efeitos. O teste de Begg é utilizado quando os efeitos possuem distribuição assimétrica, por exemplo, em metanálises de proporções.

Também é possível estimar a presença desse viés em bases de registros de estudos, pela busca daqueles que não vieram a ser publicados. Não condicione-se aos resultados estatísticos.

É necessário fazer uma apreciação crítica do processo de busca realizado, em relação aos termos, fontes de informação e idioma. Tais elementos devem ser apropriados para a pergunta da revisão sistemática.

Além do viés de publicação, as revisões sistemáticas são passíveis de interferência por outros tipos de vieses, entre eles o viés de tempo, viés de múltipla publicação, viés de localização, viés de citação, viés de linguagem e viés de desfecho. Uma ferramenta interessante para a consulta dos possíveis vieses a que um estudo está sujeito é o Catálogo de Vieses de Oxford (oxfordbrazilebm.com/index.php/catalogo-de-vieses/).

Exclusão de outliers

A exclusão de outliers (estudos atípicos) consiste na retirada da análise de estudos destoantes, responsáveis pela heterogeneidade observada. A retirada desses estudos idealmente deve ser guiada por uma justificativa de cunho clínico ou metodológico como, por exemplo, ter avaliado uma população com características diferentes dos demais estudos ou ter utilizado diferentes critérios de avaliação para o desfecho de interesse.

Etapa 7: avaliação da qualidade das evidências

Uma vez concluída a avaliação crítica do estudo e a síntese dos resultados encontrados, sugere-se avaliar a qualidade da evidência para cada desfecho priorizado na revisão sistemática. Esta etapa inclui o julgamento dos resultados de cada desfecho quanto à presença de fatores que diminuem ou aumentam a confiança na evidência. A qualidade ou o nível da evidência diz respeito ao grau de confiança ou certeza que se pode ter em uma determinada estimativa de efeito. Mas, por que classificar as evidências?

As evidências geradas pela revisão sistemática são usadas para a tomada de decisão. Nesse sentido, apesar de não necessariamente apresentarem recomendações de ações em sua conclusão, é comum que os resultados de uma revisão sistemática sejam utilizados na construção de diretrizes clínicas e políticas públicas. Apenas a hierarquização das evidências de acordo com o delineamento dos estudos não é suficiente para determinar a qualidade da evidência.

Para isso, é de grande valor ter em mãos alguma classificação que indique de forma rápida e prática o nível de confiança das evidências. Esse nível de evidência precisa ser avaliado considerando fatores individuais dos estudos primários e características da combinação deles. Apesar de bastante disseminada em décadas passadas, a hierarquização pelo delineamento das pesquisas, ou seu julgamento metodológico individual, é insuficiente para classificar de forma plena a qualidade de todo o conjunto de evidências.

Da mesma forma, a classificação em diversos níveis de evidência, como os clássicos níveis de evidência de Oxford (classificando a evidência em nível 1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 4 e 5 de acordo com o delineamento e presença de vieses), e a criação de formatos próprios e pouco difundidos de classificação pode tornar ainda mais complexa a interpretação dos leitores do que seria o resultado da classificação da evidência.

Sistema GRADE

O sistema GRADE (www.gradeworkinggroup.org) é uma proposta sensível e transparente de classificação das evidências e da força de recomendação desenvolvido por um grupo de metodologistas desde o ano 2000. O GRADE tem o objetivo de avaliar de forma clara a qualidade da evidência (grau de confiança) para cada desfecho e formular uma recomendação (contra ou a favor). Diferentemente dos instrumentos vistos até agora, como o *Risk of Bias* da Cochrane, o GRADE é aplicado a um corpo de evidências, não a estudos individuais. O GRADE tenta avaliar o conjunto da evidência para cada questão de pesquisa proposta (PICOS), considerando, além do risco de viés, outros domínios.

A abordagem GRADE é bastante difundida mundialmente, endossada por várias organizações internacionais especializadas em revisões sistemáticas e protocolos clínicos, entre elas a Organização Mundial da Saúde (OMS), o *National Institute for Health and Clinical Excellence* (NICE), o *Centers for Disease Control And Prevention* (CDC) e a colaboração Cochrane.

Classificação dos desfechos

No sistema GRADE, avalia-se o nível das evidências disponíveis para cada desfecho individual. Assim, os desfechos

incluídos devem ser clinicamente relevantes e importantes para os pacientes, e não pela informação disponível.

Essa classificação de desfechos deve preceder a revisão da evidência. Conforme proposto pelo GRADE, uma escala de 9 pontos pode ser utilizada para classificar os desfechos em três categorias: críticos (7-9 pontos), importantes (4-6 pontos) e importância limitada (1-3).

Esse processo precede o julgamento da qualidade da evidência. O desfecho continuará a ser o mais importante independentemente do nível de evidência que lhe for atribuído. Ou seja, mesmo que seja avaliado como baixa evidência, esse continuará a ser o mais importante para decisão.

Domínios do sistema GRADE

Os seguintes fatores são considerados para determinar o nível da evidência: delineamento do estudo, limitações metodológicas (risco de viés), inconsistência, evidência indireta, imprecisão, viés de publicação, magnitude de efeito, gradiente dose-resposta, fatores de confusão residuais. Comparado aos demais sistemas utilizados para a classificação das evidências, o GRADE é o mais abrangente, considerando uma gama maior de informações.

Níveis de evidência

O nível de evidência representa a confiança na informação utilizada. No sistema GRADE, a avaliação da qualidade da evidência é realizada para cada desfecho analisado utilizando o conjunto disponível de evidência (Quadro 7). Como uma das interpretações práticas, esse nível de evidência possui uma relação direta do quão confiantes ou certos estamos de que os achados obtidos no conjunto de evidências não seriam alterados significativamente por novos estudos.

É recomendada a avaliação da evidência da revisão sistemática pelo sistema GRADE, em especial se for uma revisão de intervenção. O sistema é utilizado principalmente para avaliar a qualidade da evidência para intervenções, sendo mais simples e difundido para esse uso. Contudo, há extensões que permitem a sua utilização para outros propósitos, como testes diagnósticos

e avaliação prognóstica. Também pode ser utilizado para revisões sistemáticas qualitativas, apesar de ser pouco comum.

Quadro 7. Nível de evidência do sistema GRADE. Fonte: Sistema GRADE e Ministério da Saúde, Brasil.

Nível de evidência	Definição	Implicações	Exemplos
Alto ⊗⊗⊗⊗	Há forte confiança de que o verdadeiro efeito esteja próximo daquele estimado	É improvável que trabalhos adicionais modifiquem a confiança na estimativa do efeito	Ensaio clínico randomizado sem limitações graves ou estudos observacionais bem executados, com efeito muito grande
Moderado ⊗⊗⊗○	Há confiança moderada no efeito estimado	Trabalhos futuros poderão modificar a estimativa	Ensaio clínico randomizado com limitações graves ou estudos observacionais bem executados, com efeito grande
Baixo ⊗⊗○○	A confiança no efeito é limitada	Trabalhos futuros provavelmente terão um impacto importante em nossa confiança na estimativa de efeito	Ensaio clínico randomizado com limitações muito graves ou estudos observacionais sem forças especiais ou limitações importantes
Muito baixo ⊗○○○	A confiança na estimativa de efeito é muito limitada. Há importante grau de incerteza nos achados	Qualquer estimativa de efeito é incerta	Ensaio clínico randomizado com limitações muito graves e resultados inconsistentes ou estudos observacionais com graves limitações ou observações clínicas não sistemáticas (relatos de casos ou séries de casos)

Na avaliação da evidência, o delineamento dos estudos é o primeiro aspecto a ser considerado para qualificar as evidências. Inicialmente, evidências provenientes de estudos randomizados são considerados como de alta qualidade (sendo atribuídos quatro pontos) e de estudos observacionais como de baixa qualidade (sendo atribuídos dois pontos).

Após, avalia-se a presença ou não de fatores que podem reduzir ou aumentar a confiança nas estimativas de efeito. Se uma evidência é de alta certeza, é improvável que novas pesquisas produzam mudanças substanciais nas estimativas de efeito. Por outro lado, se a certeza de um conjunto de evidências é muito baixa, futuros estudos com delineamentos mais apropriados poderão confirmar ou refutar os achados encontrados.

Primeiramente, o nível das evidências pode ser rebaixado em um ou até dois níveis se houver limitações metodológicas no delineamento e condução dos estudos, o que sugere alta probabilidade de viés. Essa aferição é feita na avaliação crítica dos estudos selecionados. Pode-se fazer uso de ferramentas de avaliação crítica de estudo primário, como o instrumento *Risk of Bias* da Cochrane, para se chegar a esse resultado.

A presença de inconsistência (heterogeneidade) também diminui nossa confiança nas estimativas. Trata-se da situação onde os estudos obtêm estimativas de efeito que diferem muito entre si. Esse fato pode ocorrer quando os resultados mudam conforme a população (por exemplo, gravidade da doença), a intervenção (por exemplo, doses baixas *versus* doses altas de um medicamento) ou o desfecho (por exemplo, diferentes definições ou aferições do desfecho entre os estudos). Para isso, pode-se fazer uso de indicadores da presença de heterogeneidade, como a estimativa do I^2 em uma metanálise. Quando há forte indício de heterogeneidade (por exemplo, $I^2 > 75\%$), é fundamental que se procure identificar as suas causas mais prováveis e, na hipótese dessa heterogeneidade não ser devidamente explicada ou mitigada, rebaixar a confiança nas estimativas.

A presença de evidência indireta ocorre quando existem diferenças substanciais entre as evidências disponíveis e a pergunta de pesquisa. Semelhante ao critério de inconsistência, devemos nos preocupar com o impacto de diferenças na população, intervenções, comparadores ou nos desfechos, contudo, não entre os estudos combinados, mas entre os estudos e a pergunta de pesquisa.

A imprecisão das estimativas de efeito é comumente identificada com a grande amplitude dos ICs. Normalmente, esse fenômeno ocorre quando existe um pequeno tamanho de amostra

ou pequeno número de eventos clínicos. Adicionalmente, pode-se fazer uso do conceito do tamanho ótimo da informação (*Optimal Information Size, OIS*), que reflete qual deveria ser o tamanho mínimo de uma amostra para se determinar com segurança um efeito observado. Caso o conjunto de evidências esteja bastante divergente desse tamanho mínimo, podemos questionar o seu nível de imprecisão.

A presença de viés de publicação seria o último fator a ser considerado no rebaixamento do nível das evidências. Sobretudo na situação de uma metanálise fortemente influenciada pelo resultado de estudos pequenos, podemos estar diante de um viés de publicação. É importante considerar se existe uma suspeita de que apenas os estudos positivos (favoráveis à intervenção ou estatisticamente significativos) foram considerados. Alguns fatores podem nortear indícios da presença ou não desse viés, como o fato de que estudos de maior porte mostraram pouco ou nenhum efeito, enquanto estudos pequenos mostraram benefícios acentuados. Idealmente, em revisões sistemáticas, esse risco de viés é investigado por análise qualitativa e mediante aplicação de testes estatísticos, como o *funnel plot*.

Em seguida, avalia-se a possibilidade de aumentar a evidência a desfechos provenientes de estudos observacionais que não tiveram sua qualidade reduzida pelos fatores que diminuem a confiança nas evidências. Desfechos provenientes de ensaios clínicos que não apresentaram preocupações graves na etapa anterior continuam com qualidade da evidência alta e também não se aplicariam a essa avaliação de fatores que elevam a confiança nas evidências.

Primeiramente, a qualidade da evidência pode ser elevada quando a magnitude do efeito for muito grande (por exemplo, $RR > 2$ ou $RR < 0,5$). Isso porque mesmo que os estudos observacionais não consigam afastar completamente a influência dos fatores de confusão, é pouco provável que toda a parcela de efeito observada seja proveniente de uma associação espúria (falsa) na situação de um efeito de grande magnitude. Outro fator que aumenta a confiança na relação de uma causa e efeito de uma intervenção é a presença de evidências de um gradiente dose-resposta. Esse critério diz respeito à situação onde observa-se um aumento do nível de efeitos com o aumento do nível de exposição.

Por fim, também é possível avaliar se existe alguma influência positiva dos fatores de confusão. Tal avaliação, de certa forma mais complexa e menos frequente que os demais fatores, diz respeito à situação em que a presença de fatores de confusão estaria favorecendo nossa confiança na existência ou ausência de efeito de intervenção. Por exemplo, na hipótese de se observar um efeito clinicamente relevante de uma intervenção nas evidências disponíveis, o nível de evidência poderia ser elevado caso se apresente algum fator de confusão que diminua o efeito da intervenção. Da mesma forma, se um fator de confusão favorece o efeito de uma intervenção e, mesmo assim, as evidências disponíveis não identificam estimativas clinicamente significantes, haveria mais confiança de que não existe um efeito daquela intervenção.

Apresentação dos dados

Em revisões sistemáticas, a avaliação GRADE geralmente é apresentada através de uma tabela sumária de resultados (*summary of findings*, SoF). Nessas tabelas, são apresentados o delineamento e o julgamento sobre cada aspecto do processo de avaliação das evidências, bem como a referência dos motivos que levaram à elevação ou rebaixamento do nível das evidências. Essa medida é importante para aumentar a transparência do processo e permitir sua adaptação aos diferentes contextos de decisão.

Existem plataformas para a formatação dessas tabelas, como o *GRADE Profile Guideline Development Tool* (GRADEpro GDT) (gradepro.org), que permite gerenciar todo o processo de avaliação das evidências e exportar as tabelas SoF em variados formatos.

Força de recomendação

Quanto à força de recomendação, o sistema GRADE também apresenta uma proposta de classificação. Ao considerar o nível de evidência e outros fatores importantes para a decisão, como os riscos e custos associados à intervenção, pode-se enquadrar uma recomendação como forte a favor, forte contra, fraca a favor ou fraca contra. Forte, onde a maioria dos indivíduos com indicação deve receber a intervenção, ou fraca (também chamada de condicional), onde o balanço de riscos e benefícios da intervenção pode ser incerto para uma parcela dos indivíduos, podendo ser a favor ou contra realizar a intervenção proposta.

Protocolos clínicos, pareceres técnico-científicos e outros documentos que envolvam a tomada de decisão em contextos definidos devem informar a força dessa recomendação além de classificar a evidência. Portanto, a força de recomendação envolve a discussão de domínios que transcendem a qualidade da evidência, como custos, custo-efetividade e preferências dos usuários. Essa discussão é variável de acordo com o cenário no qual quer se propor uma recomendação. Revisões sistemáticas, contudo, não necessariamente lidam com os contextos assistenciais ou formulam recomendações e, portanto, se limitam a apresentar apenas o nível de evidências em suas conclusões.

Portanto, a aplicação do GRADE pode ser resumida em quatro etapas:

- 1) consideração do tipo de evidência (se ensaio clínico randomizado ou estudo observacional);
- 2) avaliação dos fatores que afetam a qualidade da evidência, podendo reduzir (limitações metodológicas, inconsistências, evidência indireta, imprecisão, viés de publicação e outros fatores) ou elevar (tamanho de efeito grande, gradiente dose-resposta, fatores de confusão residuais) a qualidade da evidência;
- 3) soma dos pontos e inferência sobre a qualidade da evidência e;
- 4) formulação da força da recomendação, contra ou a favor, fraca ou forte.

O GRADE é um sistema estruturado, mas os julgamentos são susceptíveis à subjetividade. O GRADE propicia transparência, considerando a qualidade da evidência por desfecho e permite uma recomendação explícita.

Avaliação de evidências qualitativas

Para avaliação de evidências qualitativas há o *Confidence in the Evidence from Reviews of Qualitative Research* (GRADE CERQual) (www.cerqual.org), que envolve um processo criterioso e transparente de avaliação das evidências. É uma complementação do sistema GRADE que permite descrever o quão confiantes estamos sobre os achados de sínteses de evidências qualitativas. A pesquisa qualitativa busca analisar de forma mais ampla

as percepções e experiências dos indivíduos, incluindo suas perspectivas sobre saúde, serviços e políticas públicas. Por ter características e objetivos não necessariamente consistentes com a pesquisa quantitativa, as evidências provenientes da pesquisa qualitativa passam por uma avaliação específica do sistema GRADE. Mais especificamente, o nível de evidência pelo GRADE CERQual nos comunica até que ponto os achados de uma revisão seriam substancialmente diferentes do fenômeno de interesse.

Etapa 8: redação dos resultados

Nesta etapa, questiona-se se: os resultados foram interpretados com bom senso e com devida consideração aos aspectos mais amplos do problema? Ou seja, é possível generalizar as evidências e aplicar as recomendações geradas pela revisão sistemática (validade externa)?

A redação da revisão sistemática prevê a organização do texto final seguindo os guias de redação recomendados (Quadro 2), bem como a preparação de materiais que ficarão acessíveis ao público, como bancos de dados. Uma dica importante é ter familiaridade com as diretrizes de redação de revisão sistemática. Apesar da escrita da revisão sistemática ser uma das etapas finais, na prática, ela acontece no decorrer de todo o processo. Alguns ingredientes são importantes para escrever um bom artigo, independentemente de ser ou não uma revisão sistemática.

Escrever definitivamente é uma tarefa difícil. Escrever com simplicidade e clareza é ainda mais difícil. Quanto mais o autor se empenhar, menos demandará tempo ao leitor para entender o que foi feito e maior audiência o manuscrito alcançará. Persistência e disciplina são qualidades que facilitam a aprendizagem e possibilitam favorecerem o aprimoramento da escrita.

No entanto, na confecção de um artigo científico, espera-se uma escrita científica, que tem o compromisso de ser verdadeira. Falsificação de dados e argumentações exageradas tornam-se um desserviço à ciência e à sociedade. O plágio e os conflitos de interesse devem ser observados e evitados. Redigir com imparcialidade é uma qualidade essencial para um autor honesto, objetivo e equilibrado para interpretar os achados encontrados.

O leitor também espera que o autor do artigo conheça cientificamente o tema e saiba informar sobre o assunto. O conhecimento científico está disponível nos canais apropriados de divulgação da ciência, como destaque aos artigos originais, mas também na formação e capacitação formal da equipe que o desenvolve. Além de assegurar conhecimento aos autores, é estratégico circular o manuscrito aos pesquisadores mais experientes quando se alcança uma versão mais acabada do texto. Na ausência de ter com quem contar nessa etapa, pode-se lançar mão da pré-publicação (*preprints*) ou fóruns científicos.

Outro aspecto importante é o espírito científico, que se refere ao rigor e ao ceticismo organizado. A ciência se constrói um passo por vez, da mesma forma que uma revisão sistemática é sistemática. O rigor implica em incorporar um hábito que considera que novas hipóteses, ou mudanças de paradigmas vigentes, precisam passar pelo método científico. O ceticismo organizado é o hábito de cientistas resistirem a aceitar teorias e descobertas como definitivamente provadas.

O planejamento da escrita ajuda na organização das ideias. Em revisões sistemáticas, a redação do protocolo facilita o processo. Dispor de um bom registro do que foi realizado em cada etapa auxilia muito a escrita do manuscrito. No decorrer de uma revisão sistemática, é comum identificar evidências que podem nortear a interpretação dos achados.

E quando ocorre o bloqueio criativo? Quando estamos chegando ao fim de um trabalho tão laborioso como uma revisão sistemática bem conduzida, é normal e esperado que o cansaço dê sinais do seu peso. Nessa hora, vale a pena dar uma pausa, se inspirar e reenergizar para, então, com inspiração renovada, escrever um texto que tenha a clareza como prioridade. O distanciamento e desapego do texto fazem parte do processo de escrita do manuscrito. Pedir (e considerar) sugestões de colegas que podem revisar o texto, incluindo pessoas que não são da área, pode ajudar bastante na clareza. Passado esse tempo em que os colegas gentilmente se dedicaram a melhorar o trabalho, ocorre o distanciamento do texto e vale a pena uma nova leitura para identificar pontos que precisam de maior clareza.

Por fim, revisões sistemáticas devem ter conclusões claras e definidas, sem possibilidade de confusão ao leitor. A conclusão resume de maneira simples e direta o que se pode afirmar a partir dos achados da revisão sistemática e deve responder aos objetivos do estudo. É necessário evitar afirmações que não tenham sustentação nos resultados. Inspeccionar essa seção em artigos previamente publicados facilitará o processo. Na medida do possível, a conclusão deve realçar no que a revisão sistemática se difere das demais e o que se tem de novo frente ao que se já conhecia sobre o assunto. Lembre-se: a conclusão deve responder à pergunta específica da revisão sistemática.

Etapa 9: publicação e atualização dos resultados

Por fim, a publicação inclui os esforços de ajustes às regras da revista escolhida, submissão, resposta e ajustes referentes aos pareceres recebidos, revisão e aprovação da versão final. O modo mais conveniente de comunicar os achados de uma revisão sistemática é por meio de artigos científicos. Em um ciclo perfeito, o editor do periódico identificará a relevância da investigação e encaminhará à revisão por pares (*peer review*), em que dois ou mais pesquisadores avaliarão o texto e irão sugerir pontos a serem revistos pelos autores, que têm oportunidade de ajustar o manuscrito e encaminhar para apreciação. Em caso de recusa, esse ciclo é repetido até o artigo ser aceito, o que pode levar em torno de 4-12 meses ou mais.

Por mais que a pesquisa não confirme hipóteses prévias, os resultados devem ser disseminados, mesmo quando negativos, até para evitar pesquisas redundantes em algo que não funciona, ou para permitir o aprimoramento dos métodos em eventual replicação da pesquisa.

É necessário conhecer o periódico em que se pretende submeter o artigo, avaliando se é a escolha certa para o artigo. O periódico deve ter boa avaliação, ter um bom fator de impacto e ser um periódico indexado. O escopo da revista também deve ser considerado. É comum, por exemplo, a revista ser da área, mas não aceitar determinados delineamentos, como revisões sistemáticas. Cada revista tem suas regras e é estratégico familiarizar-se com essas normas e atendê-las minuciosamente.

Custos com a publicação devem ser observados, pois alguns periódicos fazem cobranças pela publicação na modalidade *Open Access*, por exemplo. Gastos eventuais também podem ocorrer, como exigência de revisão idiomática certificada, como o *Proof-Reading-Service* (www.proof-reading-service.com/).

Algumas ferramentas podem auxiliar na seleção do periódico, como o *Edans Journal Selector* (en-author-services.edanzgroup.com/journal-selector) ou o seletor de jornais de editoras científicas, como da Elsevier (journalfinder.elsevier.com/) ou da Springer (journalsuggester.springer.com/). Nessas ferramentas é possível inserir o título, resumo e palavras-chave do manuscrito e receber uma lista de sugestão de revistas relacionadas ao tema do estudo.

É importante conhecer o estilo da redação da revista para onde a revisão sistemática será submetida. Erros ortográficos, semânticos e gramaticais podem distorcer um bom trabalho científico e reduzir a apreciação de um manuscrito após submissão, por mais que se trate de resultado robusto e inovador. Dispor de boas ilustrações e tabelas enaltece o trabalho e facilita o entendimento. Vale a pena despende um bom tempo nessa elaboração até que se alcance ilustrações verdadeiramente explicativas, elucidativas e de boa qualidade.

Entre os pontos importantes deste processo, destacam-se o seguimento do o estilo das referências, da forma de citação no texto (por exemplo, sobrescrito, parênteses, espaçamento, pontuação), das regras gerais de formatação (uso de letras maiúsculas nos títulos, subtítulos e texto, espaçamento e pontuação do texto, indicadores para notas de rodapé), além de informações mínimas exigidas para títulos de ilustrações, formato de arquivo para as figuras (JPG, TIF, PNG), resolução das figuras e tamanho aceitável para tabelas (limites de linhas e/ou colunas).

Além dessas etapas, espera-se que o grupo se comprometa com a atualização da revisão após a publicação. A atualização somente deve ser considerada e realizada se for possível melhorar a certeza em torno das conclusões da revisão "original", seja para modificar os resultados, seja para modificar a credibilidade dos achados. Por exemplo, uma revisão sistemática em que não foi possível realizar metanálise e, verificando novos estudos, passasse a poder realizá-la, é um bom argumento para atualizar a revisão.

Atualizações desnecessárias devem ser evitadas, pois apenas adicionam volume ao número de estudos já existente, podendo gerar conclusões conflitantes.

Referências

Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Meth* [Internet]. 2005 [2020 Set 11];8(1):19-32. Disponível: <https://www.york.ac.uk/inst/spru/pubs/pdf/Scopingstudies.pdf>

Bonita R.; Beaglehole, R.; Kjellström, T. *Epidemiologia básica*. 2. ed. São Paulo: Santos, 2010-2019. 213 p.

Brasil, Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Diretrizes metodológicas: elaboração de revisão sistemática e metanálise de ensaios clínicos randomizados. Série A. Normas e Manuais Técnicos. Brasília: Editora do Ministério da Saúde, 2012.

Brasil, Ministério da Saúde. Rede Brasileira de Avaliação de Tecnologias em Saúde. Diretrizes metodológicas: Sistema GRADE. Manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde. Brasília -DF, 2014.

Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York (UK): University of York; 2009. Available from: <http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm>.

Cipriani A, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*. 2018;391(10128):1357-1366.

Greenhalgh T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook

Lefebvre C, Glanville J, Wieland LS, Cole B, Weightman AL. Methodological developments in searching for studies for systematic reviews: past, present and future? *Syst Rev*. 2013 Sep 25;2:78. doi: 10.1186/2046-4053-2-78.

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1. doi: 10.1186/2046-4053-4-1.

Rodrigue MGA, Krass-Silva L, Martins ACM. Metanálise de ensaios clínicos de intervenção familiar na condição esquizofrenia. *Cad. Saúde Pública* 2008;24(10): 2203-2218.

Estudos secundários: diretriz clínica e análise econômica

*Fernanda Dapper Machado
Fernanda Vargas Ferreira
Edison Capp
Otto Henrique Nienov*

Diretrizes Clínicas

Diretrizes clínicas são recomendações elaboradas de forma sistemática para auxiliar a tomada de decisão dos profissionais da saúde e do paciente sobre os cuidados de saúde mais apropriados em circunstâncias clínicas específicas. As diretrizes clínicas podem ser produzidas por sociedades especializadas para orientar o manejo ideal (diretrizes terapêuticas) ou por governos na regulação do acesso à insumos e serviços de saúde (protocolos clínicos). As revisões sistemáticas são utilizadas para a formulação destas diretrizes clínicas.

As diretrizes clínicas constituem documentos resumidos, baseados em evidências científicas organizadas sistematicamente, que consideram todos os grupos de pacientes (de baixo e de alto risco), as opções de manejo e os possíveis desfechos relevantes (mortalidade, morbidade, qualidade de vida, toxicidade, efeitos adversos, inconveniência, carga psicológica e custos para o paciente ou para a sociedade). Em uma boa diretriz, as evidências científicas são organizadas sistematicamente, o grupo que desenvolve a diretriz inclui representantes de todas as disciplinas relevantes, inclusive pacientes, e as recomendações são explicitamente vinculadas às evidências das quais derivaram. As diretrizes para diferentes estabelecimentos e diferentes propósitos podem ser embasadas nas mesmas evidências, mas resultar em recomendações distintas.

Protocolos são instruções sobre o que fazer em circunstâncias específicas. São semelhantes às diretrizes, mas deixam menos espaço para raciocínio individual e, com frequência, são produzidos para pessoal menos experiente ou para uso em situações nas quais as eventualidades são previsíveis.

No âmbito do Sistema Único de Saúde (SUS) tais documentos têm caráter normativo e orientativo e são denominados Protocolos Clínicos e Diretrizes Terapêuticas (PCDT), desenvolvidas pelo Ministério da Saúde (www.saude.gov.br/protocolos-e-diretrizes). Os PCDT são documentos que estabelecem critérios para o diagnóstico da doença ou do agravo à saúde, o tratamento preconizado (com os medicamentos e demais produtos apropriados), as posologias recomendadas, os mecanismos de controle clínico, e o acompanhamento e a verificação dos resultados terapêuticos, a serem seguidos pelos gestores do SUS. São baseados em evidência científica e consideram critérios de eficácia, segurança, efetividade e custo-efetividade das tecnologias recomendadas.

Além do PCDT, há Diretrizes Diagnósticas e Terapêuticas (DDT) na área de Oncologia, baseados em evidência científica que visam nortear as melhores condutas na área oncológica. A principal diferença em relação aos PCDT é que, por conta do sistema diferenciado de financiamento dos procedimentos e tratamentos em oncologia, este documento não se restringe às tecnologias incorporadas no SUS, mas sim ao que pode ser oferecido a este paciente, considerando que o financiamento é repassado como procedimento para o atendimento aos centros de atenção e a autonomia destes na escolha da melhor opção para cada situação clínica. Os Protocolos de Uso são documentos normativos de escopo mais estrito, que estabelecem critérios, parâmetros e padrões para a utilização de uma tecnologia específica em determinada doença ou condição. Protocolos deixam menos espaço para raciocínio individual. As Diretrizes Nacionais Brasileiras são documentos norteadores das melhores práticas a serem seguidas por profissionais de saúde e gestores, sejam do setor público ou privado da saúde. As Linhas de Cuidados apresentam a organização do sistema de saúde para garantir um cuidado integrado e continuado, com o objetivo de atender às necessidades de saúde do usuário do SUS em sua integralidade.

Além das diretrizes nacionais, pode-se consultar o *National Guideline Clearinghouse* (www.guideline.gov/), que é uma iniciativa da *Agency for Healthcare Research and Quality* (AHRQ) dos Estados Unidos, que reúne diferentes diretrizes, desde níveis de evidência até às recomendações. Todas as diretrizes são atualizadas e revisadas a cada 5 anos. O *National Institute for Health and Care Excellence* (NICE, cks.nice.org.uk/), do Reino Unido, reúne diretrizes direcionadas para quem atua em atenção primária à saúde. O TRIP (*Turning Research into Practice*, www.tripdatabase.com) é uma ferramenta de busca integrada para encontrar diretrizes, incluindo a NICE e o *National Guideline Clearinghouse*.

Mas, por que diretriz clínica?

Os objetivos das diretrizes clínicas incluem:

- 1) tornar os padrões baseados em evidências explícitos e acessíveis;
- 2) identificar estratégias efetivas;
- 3) tornar a tomada de decisões mais fácil e objetiva;
- 4) padronizar a qualidade do serviço;
- 5) educar pacientes e profissionais sobre a “melhor prática corrente”;
- 6) fornecer um guia para avaliar o desempenho profissional;
- 7) reduzir eventos adversos (segurança do paciente);
- 8) melhorar a relação custo-efetividade dos serviços de saúde;
- 9) servir como ferramenta para controle externo;
- 10) identificar tecnologias obsoletas que devem ser descartadas e;
- 11) delinear a divisão de trabalho.

Os aumentos na demanda de cuidados com o envelhecimento e aumento de morbidades da população e, conseqüentemente, de custos dos serviços de saúde, exige o desenvolvimento contínuo de diretrizes clínicas.

Evento adverso

O evento adverso é qualquer evento (sinal, sintoma ou doença), não desejável, que ocorre com o participante de pesquisa no decorrer do estudo. Um evento adverso é considerado grave quando coloca a vida do paciente em risco, podendo: levar a óbito; causar deficiência e/ou dano permanente; implicar na necessidade de intervenção médica, cirúrgica ou hospitalização para prevenir comprometimento permanente; levar a perturbação ou risco fetal, morte fetal ou anomalia congênita ou; qualquer transmissão de agente infeccioso. O evento adverso inesperado corresponde a qualquer ocorrência que surja no decorrer da pesquisa, cuja natureza ou severidade não é consistente com as informações aplicáveis ou conhecidas do produto (sem descrição na bula, monografia, brochura ou protocolo de estudo).

Por que seguir as diretrizes clínicas?

Alguns profissionais da saúde demonstram certa relutância para seguir as diretrizes clínicas, muitas vezes por desconhecimento da diretriz, falta de subsídios claros para a sua implementação, por não acreditar que a diretriz leva a um desfecho desejado, inércia ou falta de motivação, além de barreiras externas, como falta de recursos financeiros e preferências dos pacientes.

Para prevenir problemas na elaboração de diretrizes, convém ressaltar que essas não devem refletir a opinião de um determinado especialista ou a prática padronizada vigente. Recomenda-se que haja uma atualização constante na sua elaboração, que haja um diálogo interprofissional e que se considere a realidade local da Atenção à Saúde nos diferentes níveis (primária, secundária e terciária). Ademais, essas devem propiciar a implementação de novas evidências de pesquisa.

Desenvolvimento de diretrizes clínicas

É recomendável um processo transparente e sensível na identificação, seleção e combinação da evidência disponível, com definição da pergunta de pesquisa e prioridade de desfechos. Para sua elaboração, esses documentos devem, em síntese, identificar revisões sistemáticas de boa qualidade (avaliar o risco de viés das revisões sistemáticas), avaliar a confiança dos resultados

encontrados (pelo sistema GRADE), formular as recomendações e direção destas (contra ou a favor) e deliberar sobre a força das recomendações (forte ou fraca/condicional).

Avaliação da qualidade da diretriz clínica

O *Appraisal of Guidelines for Research and Evaluation II* (AGREE II) (www.agreetrust.org/agree-ii/) é utilizado para avaliar a qualidade metodológicas das diretrizes clínicas, fornecendo uma estratégia de metodologia para o seu desenvolvimento, informando que e como as informações devem ser relatadas nas diretrizes clínicas. É composto por 23 itens, divididos em seis domínios:

- 1) escopo e finalidade (descreve objetivo, questão de saúde e população-alvo);
- 2) envolvimento das partes interessadas (a equipe deve incluir indivíduos de todos os grupos profissionais relevantes, procurando conhecer a opinião e preferências da população-alvo, além de definir os cenários-alvo claramente);
- 3) rigor do desenvolvimento (utilizar-se de métodos sistemáticos de busca e critérios para seleção de evidências);
- 4) clareza e apresentação (as recomendações devem ser específicas e sem ambiguidade);
- 5) aplicabilidade (a diretriz deve descrever os fatores facilitadores e as barreiras para implantação, aconselhamentos, implicações e monitoramento) e;
- 6) independência editorial (o financiador não deve exercer influência sobre o conteúdo da diretriz, além de inferir possíveis conflitos de interesse).

A ferramenta *Adaptation of clinical guidelines* (ADAPTE) (DOI: 10.1111/jep.12479) propicia uma abordagem sistemática de aproveitamento das informações disponíveis nas diretrizes previamente realizadas para a circunstância local. A ferramenta ADAPTE foi desenvolvida pelo *ADAPTE Working Group Collaboration*, é validada e procura melhorar a eficiência de diretrizes existentes, ou seja, a adaptação de diretrizes em um ambiente cultural e organizacional diferente. O objetivo é aproveitar as diretrizes existentes para melhorar a eficiência da produção e utilização de adaptações de alta qualidade. Compreende três fases:

- 1) configuração (preparação);
- 2) adaptação (definição do escopo e propósito, pesquisa e filtro, avaliação, decisão e seleção e, customização) e;
- 3) finalização (revisão externa, planejamento e acompanhamento e, produção final).

Podemos identificar três aspectos em uma boa diretriz:

- 1) conteúdo (por exemplo, se foi baseada em uma revisão sistemática abrangente e rigorosa);
- 2) processo (como a diretriz foi organizada) e;
- 3) apresentação (quão atraente é e qual o grau de facilidade para ser seguida).

Ao avaliar uma diretriz clínica, algumas questões norteadoras podem ser seguidas:

A preparação e a publicação da diretriz envolveram algum conflito de interesses significativo? Lembre-se que uma diretriz pode ser baseada em evidências, mas o problema abordado por ela terá sido elaborado por uma equipe que vê o mundo de determinada maneira. A diretriz deve mencionar as pessoas que a desenvolveram, financiaram ou apoiaram, como patrocinadores e parcerias.

A diretriz está relacionada a um tópico apropriado e menciona claramente a população-alvo ao qual se aplica? Neste aspecto, deve-se observar o objetivo principal da diretriz, incluindo o problema de saúde e os pacientes, profissionais e ambientes-alvo. Questione se o tópico é importante, se é possível implementar, qual o benefício e o investimento para sua implementação.

O grupo de desenvolvimento da diretriz inclui pelo menos: um especialista na área do tópico, um especialista nos métodos de pesquisa secundária e uma pessoa afetada pelo problema? Se uma diretriz clínica foi inteiramente preparada por um grupo, você deve ter um olhar mais crítico, pois a diretriz pode estar refletindo a opinião do especialista ou pode ser percebida como um ato político, não refletindo as necessidades da população-alvo.

Os raciocínios subjetivos do grupo de trabalho que desenvolveu a diretriz foram explicitados e justificados? Desenvolver

uma diretriz não é apenas um processo técnico de descoberta, análise e transformação de evidências em recomendações. As recomendações também exigem raciocínio. Devem levar em conta os princípios éticos, preferências, culturas e aspirações.

Todos os dados relevantes foram investigados e rigorosamente avaliados? Questione se as evidências estão apoiadas em estudos de pesquisas primárias de alta qualidade ou se baseiam na opinião de especialistas. A literatura disponível deve ser criteriosamente analisada e não ser uma afirmativa prática de um grupo de especialistas. Todos os artigos incluídos devem ser metodologicamente avaliados pelo sistema GRADE.

As evidências foram adequadamente sintetizadas e as conclusões da diretriz refletem os dados sobre os quais foram embasadas? Verifique como os diferentes estudos foram combinados e como e quando as evidências foram coletadas, selecionadas e sintetizadas.

A diretriz aborda as variações na prática clínica e em outras áreas controversas? Questione se as opções da prática clínica foram consideradas na formulação da diretriz e se as diferentes áreas foram respeitadas.

A diretriz é clinicamente relevante, abrangente e flexível? Examine a perceptiva a partir da qual a diretriz foi redigida. Se elaborada da perspectiva do paciente, veja quais circunstâncias foram consideradas, se a diretriz abrange todas ou a maioria das eventualidades clínicas. Afinal, são os pacientes "atípicos" que motivam os profissionais a procurar por diretrizes. A flexibilidade é uma consideração particularmente importante para os órgãos nacionais e regionais que desenvolvem diretrizes. A apropriação das diretrizes pelas próprias pessoas que irão usá-las é fundamental para a sua real utilização.

A diretriz leva em conta o que é aceitável, financeiramente acessível e possível na prática para o paciente? Examine o tipo e a magnitude dos benefícios, dos danos e dos custos esperados para os pacientes a partir da implementação da diretriz, se os desfechos significativos em termos econômicos e de saúde foram considerados e se incorpora as necessidades e prioridades dos pacientes no seu desenvolvimento.

A diretriz inclui recomendações para sua própria divulgação, implementação e revisão periódica? Uma boa diretriz deve sugerir métodos para maximizar o seu uso e assumir a responsabilidade de ser revisada e atualizada periodicamente em suas recomendações.

Análise econômica em saúde

A análise econômica envolve o uso de técnicas analíticas para definir opções na alocação de recursos, comparando duas ou mais tecnologias, métodos, tratamentos, etc., em termos de custos e consequências (impacto monetário) para o indivíduo e seu meio social. Tem o objetivo principal de fornecer suporte para uma tomada de decisão consciente, fazendo com que a alocação do recurso seja eficiente (otimização).

Quando nos dirigimos a análise econômica, podemos estar avaliando diferentes alternativas de tratamento, tecnologias, medicamentos, instrumentos, métodos ou terapia, em relação a uma pergunta central. A pergunta central é o que define o tipo de análise econômica a ser realizada, se de custo-minimização, custo-benefício, custo-efetividade, custo-utilidade ou impacto orçamentário.

Análise de custo-minimização

A análise de custo-minimização não utiliza medida de desfecho. É utilizada quando o efeito de ambas as intervenções é sabidamente idêntico, ou seja, são equivalentes, mas diferem em termos de custo. Compara-se unicamente em relação aos custos com a premissa de equivalência das alternativas em termos de desfechos clínicos (eficácia e segurança), sociais, éticos e legais. Tem como vantagem a simplicidade, de ser útil na comparação de diferentes doses e vias de administração, com a limitação de não avaliar desfechos ou os resultados clínicos. Depende de uma evidência contundente sobre a equivalência das alternativas de estudo.

Análise de custo-efetividade

Na análise de custo-efetividade, as medidas de desfecho são unidades naturais (por exemplo, anos de vida gastos). É utilizada quando o efeito das intervenções pode ser expresso em termos de uma variável principal. Compara-se em termos de custos e resultados clínicos, identificando a alternativa em saúde mais barata que atinge

o mesmo objetivo em saúde, ou seja, quantifica o benefício clínico obtido por unidade monetária investida. O benefício é mensurado por meio de desfechos clínicos e sociais e as alternativas são avaliadas, mesmo que muito diferentes, pelo seu benefício na mesma unidade. Tem a vantagem da interpretação objetiva (custo/unidade de efetividade) e é limitada pela dependência de dados de efetividade (funcionamento da alternativa em condições reais). Na literatura, são mais frequentes dados de eficácia (funcionamento da alternativa em condições ideais).

Análise de custo-utilidade

A análise de custo-utilidade utiliza como medidas de desfecho as medidas de utilidade (por exemplo, anos de vida ajustados pela qualidade). É utilizada quando o efeito das intervenções sobre o estado de saúde tem duas ou mais dimensões importantes (por exemplo, benefícios e efeitos colaterais de fármacos). É semelhante à análise de custo-efetividade, exceto pelo fato do desfecho ser avaliado como ano de vida ganho e ajustado pela qualidade (*Quality-Adjusted Life Year, QALY*). Ou seja, o tempo de sobrevivência das pessoas ajustado pela qualidade de vida. O QALY é a medida que captura o benefício da alternativa em termos de anos de vida a mais (sobrevivência) em perfeita saúde. Por exemplo, 5 QALYs significa que a alternativa fornece 5 anos com uma qualidade de vida perfeita.

A “qualidade de vida” é quantificada pelo *status* do indivíduo, em um determinado período. Algumas avaliações também consideram estados de saúde negativos, como dor ou desconforto extremo sem possibilidade de controle terapêutico adequado, confinamento em cama, etc. O questionário SF-6D (*Short-Form 6 Dimensions*) é um questionário derivado dos itens do SF-36 (*Medical Outcomes Study 36-Item Short-Form Health Survey*) para descrever estados de saúde e gerar índices de utilidade. É composto por 6 domínios que possuem de 4 a 6 níveis, podendo gerar 18.000 estados de saúde. O sistema descritivo na língua portuguesa foi desenvolvido por pesquisadores da Universidade Federal de São Paulo e o algoritmo para cálculo dos escores foi elaborado por um grupo de pesquisa da Universidade Federal do Rio Grande do Sul. Tem a vantagem de comparação de alternativas diferentes na mesma unidade e interpretação simples (custo/DALY) e a desvantagem da dificuldade e subjetividade em capturar

quantitativamente a qualidade de vida, além da dependência entre a severidade da doença e a importância dada para uma unidade de QALY. A interpretação adequada de uma análise de custo-utilidade precisa considerar a severidade da doença ou condição de saúde.

Análise de custo-benefício

O custo-benefício usa medidas de desfechos em unidades monetárias (por exemplo, custo estimado da perda em produtividade). É utilizada quando é desejável comparar uma intervenção para um problema com uma intervenção para outro problema. Transforma os benefícios em valores monetários e os custos diretos e indiretos são incorporados. Tem a vantagem de os tomadores de decisão poderem maximizar o bem-estar da sociedade com base em um recurso finito, não necessariamente comparando duas alternativas. Apresenta dificuldade de conversão do benefício em unidades monetárias, imprecisão, subjetividade e variabilidade interpessoal.

Análise de impacto orçamentário

A análise de impacto orçamentário utiliza a medida de desfecho de impacto financeiro. É usada para a adoção ou desincorporação de uma alternativa no sistema de saúde. Prevê qual o impacto financeiro global da adoção (ou desincorporação) de uma alternativa no sistema de saúde, baseado em evidências. É uma ferramenta essencial para a tomada de decisão que permite avaliar se há recursos suficientes para sustentar uma decisão.

Elaboração de análise econômica

A elaboração de uma análise econômica em saúde pode ser dividida em nove etapas:

- 1) definição dos tratamentos, desfechos e tipo de análise econômica (define-se a questão de pesquisa, as alternativas a serem avaliadas (duas ou mais para a comparação), a população-alvo e o desfecho de interesse);
- 2) adoção de uma perspectiva de análise (seguradora, SUS, sociedade, indústria, hospital, plano de saúde);
- 3) escolha de um horizonte temporal adequado (quantidade de tempo que as alternativas serão avaliadas

para definir os benefícios e os custos associados com a nova alternativa);

4) identificação dos tipos de custos (os custos a serem incluídos dependem da perspectiva adotada e normalmente incluem custos de curto e longo prazo, diretos e indiretos);

5) adoção de uma taxa de desconto (ajuste para diferenças temporais em termos de custos, de forma que os custos futuros sejam comparáveis aos custos atuais);

6) realização da análise econômica (pode ser feita com auxílio de softwares como Excel, TReeAge, Arena e R);

7) análise de sensibilidade (avalia a robustez dos modelos frente à modificação, como variação dos parâmetros utilizados);

8) apresentação das fontes de dados e informações (deve-se utilizar fontes de informação de qualidade e compatíveis com a realidade, detalhando as fontes, critérios de escolha, limitações e vieses) e;

9) identificação da fonte de financiamento do estudo e potenciais conflitos de interesse.

Avaliação da qualidade da análise econômica

As análises econômicas também são utilizadas para tomada de decisão quanto às políticas de saúde, incorporação de novas tecnologias, implementação de programas preventivos e desenvolvimento de diretrizes assistenciais na área da saúde. Portanto, torna-se necessário avaliar e analisar criticamente estas análises. É importante considerar que existem inúmeros estudos de avaliação econômica, incluindo análises de custo de doença, estudos de custo-benefício, custo-efetividade e custo-minimização, entre outros, que são facilmente acessíveis em diversos periódicos e também publicações ou documentos técnicos de diversas instituições.

A dificuldade de se generalizar estes resultados deve-se a inúmeros fatores, entre eles o problema de se generalizar dados de ensaios clínicos (condições ideais) para a prática do dia a dia (condições reais), a dificuldade de comparabilidade de dados

econômicos ao longo do tempo, ou a dificuldade de generalização de resultados de estudos realizados em lugares diferentes em que alguns dos pressupostos do modelo analítico podem ser variáveis.

Desta forma, a avaliação metodológica das análises econômicas é necessária, para respondermos questões como: quais os principais fatores limitantes da generalização dos resultados da análise-econômica? Como interpretar os resultados de estudos conduzidos em diferentes cenários? Podemos utilizar alguns resultados, quando são baseados em dados dos sistemas de saúde de outros países? Quais as recomendações metodológicas para realização de estudos que podem ser traçadas a fim de minimizar este problema?

Existem guias que auxiliam na análise crítica de estudos de análise econômico, como o desenvolvido por Nunes da Silva et al., 2014 (DOI: 10.5123/S1679-49742016000100023). A seguir, são abordados aspectos que devem ser considerados na avaliação da análise econômica.

A análise é baseada em um estudo que responde a uma questão clínica definida com clareza sobre um assunto economicamente importante? Antes de analisar os custos, escalas de qualidade de vida ou serviços, certifique-se de que o estudo é cientificamente relevante e capaz de fornecer respostas nem tendenciosas e nem ambíguas à questão de pesquisa apresentada na introdução.

Do ponto de vista de quem os custos e os benefícios estão sendo considerados? Verifique sob qual ponto de vista foi desenvolvida a análise econômica: do paciente, das finanças públicas, da empresa farmacêutica, do plano de saúde, etc. Não existe uma análise econômica destituída de perspectiva. Não há perspectiva ideal para uma avaliação econômica, mas o artigo deve dizer claramente de quem são os custos e os benefícios que foram adicionados e excluídos. A maioria assume a perspectiva do sistema de saúde em si, embora algumas levem em conta os custos para o paciente e para a sociedade.

Quando comparadas, as intervenções demonstraram ser clinicamente efetivas? Certifique-se de que a intervenção que

apresenta menos custos não seja substancialmente menos efetiva. É importante lembrar que o sistema de saúde tem recursos limitados. Considere o impacto da implementação da intervenção a curto, médio e longo prazo.

As intervenções são adequadas e executáveis nas circunstâncias em que provavelmente serão aplicadas? Deve haver informação sobre a pertinência da intervenção para o problema de saúde em questão, além de uma descrição detalhada da alternativa e dos fatores facilitadores e as barreiras para implantação.

Qual método de análise foi usado? O tipo de avaliação econômica está relacionado ao desfecho principal escolhido no estudo: se for desfecho clínico, custo-efetividade; se for medida subjetiva (por exemplo, anos de vida ajustado pela qualidade), custo-utilidade; se for benefício monetário, custo-benefício; se não houver diferença em desfecho, custo-minimização.

Como foram medidos os custos e os benefícios? Caso as estimativas dos desfechos em saúde tenham sido baseadas em ensaios clínicos, deve haver conformidade entre o protocolo de pesquisa e a prática médica retratada no estudo. Caso as estimativas dos desfechos em saúde tenham sido baseadas em revisão sistemática, deve haver informação sobre a qualidade da evidência da revisão sistemática e dos estudos primários considerados. Caso as estimativas dos desfechos em saúde tenham sido baseadas em dados observacionais, deve haver uma justificativa para tal uso, como por exemplo, ausência de evidência de melhor qualidade.

Foi realizada uma análise de sensibilidade? A análise econômica deve considerar diferentes formas de análise e fatores para ter-se certeza da generalização (validade externa). Existem diversos fatores que interferem nos resultados de análises econômicas e dificultam a sua generalização para outros contextos, como as diferenças demográficas e epidemiológicas das doenças, a disponibilidade de recursos médicos e variabilidade da prática médica, os incentivos e as formas de remuneração dos profissionais e instituições de saúde e as diferenças no custo absoluto e relativo dos aspectos relacionados à saúde.

Referências

Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Meth* [Internet]. 2005 [2020 Set 11];8(1):19-32. Disponível: <https://www.york.ac.uk/inst/spru/pubs/pdf/Scopingstudies.pdf>

Bonita R.; Beaglehole, R.; Kjellström, T. *Epidemiologia básica*. 2. ed. São Paulo: Santos, 2010-2019. 213 p.

Brasil, Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. Diretrizes metodológicas: elaboração de revisão sistemática e metanálise de ensaios clínicos randomizados. Série A. Normas e Manuais Técnicos. Brasília: Editora do Ministério da Saúde, 2012.

Brasil, Ministério da Saúde. Rede Brasileira de Avaliação de Tecnologias em Saúde. Diretrizes metodológicas: Sistema GRADE. Manual de graduação da qualidade da evidência e força de recomendação para tomada de decisão em saúde. Brasília -DF, 2014.

Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. York (UK): University of York; 2009. Available from: <http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm>.

Cipriani A, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*. 2018;391(10128):1357-1366.

Greenhalgh T. *Como ler artigos científicos: Fundamentos da medicina baseada em evidências*. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*. Cochrane, 2019. Available from www.training.cochrane.org/handbook

Lefebvre C, Glanville J, Wieland LS, Cole B, Weightman AL. Methodological developments in searching for studies for systematic reviews: past, present and future? *Syst Rev*. 2013 Sep 25;2:78. doi: 10.1186/2046-4053-2-78.

Moherd D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1. doi: 10.1186/2046-4053-4-1.

Rodrigues MGA, Krauss-Silva L, Martins ACM. Metanálise de ensaios clínicos de intervenção familiar na condição esquizofrenia. *Cad. Saúde Pública* 2008;24(10): 2203-2218.

Estudos de intervenção: ensaios de campo, comunitário e clínico

*Laís Lima Ferreira
Nadine de Souza Ziegler
Edison Capp
Otto Henrique Nienov*

Ao contrário dos estudos observacionais, em que se obtém a evidência em um “mundo real” através da observação dos eventos (desfecho e exposição), nos estudos intervencionais o pesquisador planeja um experimento controlado para compreender o efeito (eficácia e segurança) de um tratamento ou intervenção, avaliando assim a evidência em um “mundo ideal”.

Os estudos de intervenção (ou experimentais) envolvem a condução de um ou mais tratamentos ou intervenções (medicamento, procedimento cirúrgico, diretriz clínica, programa de educação em saúde, etc.) pelo pesquisador de modo a modificar, de alguma forma, um processo de doença. Se dividem em ensaio de campo, ensaio comunitário e ensaio clínico randomizado ou não randomizado (quase-experimental). Simplificadamente, podemos diferenciá-los pela unidade de estudo: se os participantes do estudo forem pessoas saudáveis, o ensaio será de campo; se os participantes forem os próprios membros de uma comunidade específica, o ensaio será comunitário; e se os participantes forem os pacientes, ou seja, pessoas doentes em que se quer testar um medicamento ou tratamento, o ensaio será clínico.

Ensaio de campo

A população estudada nos ensaios de campo constitui-se de indivíduos livres de doença, mas sob risco de desenvolvê-la no ambiente em que se encontram (por isso, “no campo”). O principal objetivo é desenvolver e avaliar técnicas de prevenção. Por exemplo, estudos sobre vacinas (como o da vacina Salk

para prevenção da poliomielite, que envolveu 1,8 milhões de participantes, nos Estados Unidos, na década de 50), exposição a inseticidas, presença de chumbo em tintas para crianças.

A condução de ensaios de campo favorece tanto a implementação de medidas de saúde pública, como a manutenção de estudos de prevalência e de incidência da condição de saúde em estudo. A desvantagem é que requer um número grande de indivíduos, além de ser oneroso e logisticamente complicado para se conduzir e acompanhar por um longo período de tempo.

Ensaio comunitário

Nos ensaios comunitários, os participantes são os próprios membros da comunidade, ou seja, são comunidades ao invés de indivíduos. Avalia-se as doenças relacionadas às condições sociais, desde que sejam facilmente influenciadas por intervenções dirigidas ao comportamento do grupo ou do indivíduo. O objetivo principal é analisar as comunidades inteiras. As campanhas para prevenção da Síndrome da Imunodeficiência Adquirida (AIDS), fluoretação da água para prevenção de cárie e uso de inseticida no controle de vetores (como a dengue), são exemplos de ensaios comunitários.

Dentre as limitações, destaca-se a dificuldade de isolar uma comunidade devido as mudanças sociais, um pequeno número de comunidades pode ser incluído e a alocação aleatória não é muito prática. Por esse motivo, opta-se por comunidades homogêneas entre si para assegurar que o resultado observado seja explicado pela intervenção e não a diferenças inerentes às comunidades.

Ensaio clínico

No ensaio clínico, os participantes são os pacientes. O objetivo principal é avaliar o efeito de um tratamento em particular. Na sua forma mais simples, os ensaios clínicos são compostos por um grupo de interesse em que se faz uso de uma terapia ou exposição (grupo de intervenção) comparando-se com um grupo controle, no qual os indivíduos selecionados são aleatoriamente (randomizado) ou não alocados (não randomizado) e compara-se os desfechos entre os grupos.

Fases do ensaio clínico

Os ensaios clínicos são amplamente utilizados para avaliar a eficácia e segurança de novos medicamentos. Devido a sua complexidade, o desenvolvimento de um novo medicamento até sua aprovação e comercialização envolve diferentes fases de estudo. Os estudos pré-clínicos (fase 0), que antecedem os estudos clínicos, utilizam modelos *in vitro* e *in vivo* para avaliar a toxicidade e a letalidade do medicamento e a possibilidade de progressão para estudos clínicos. Posteriormente, nos estudos clínicos busca-se conhecer a eficácia, a tolerabilidade e a segurança do novo medicamento em seres humanos.

Pode-se caracterizar quatro fases:

1. Fase I: são os primeiros estudos realizados em seres humanos, não sendo considerados um ensaio clínico. O objetivo principal é avaliar a farmacocinética e farmacodinâmica do medicamento em estudo (dosagens seguras e efeitos colaterais). São estudos de curta duração (6-9 meses), dependendo da meia-vida e do prazo esperado para manifestação de efeitos colaterais previsíveis. Envolve de 20 a 100 participantes saudáveis, onde os critérios de inclusão e exclusão visam afastar a possibilidade de doenças prévias ou de limitações que possam representar um risco à integridade do participante da pesquisa.

2. Fase II: corresponde aos estudos de eficácia, tolerabilidade e segurança. Apresentam maior duração (6-24 meses) e número de participantes (100 a 1000), mas agora com a doença ou condição de saúde específica. Estes estudos podem ser realizados em um ou mais centros de pesquisa. Além disso, podem ser divididos em duas fases: fase precoce ou IIA, realizada em pequenos grupos de participantes que possuem a doença para qual o novo medicamento está sendo testado em paralelo com um grupo controle (placebo ou controle ativo), com critérios de inclusão e exclusão bastante restritivos e rigorosos, visto que tem como objetivos: prova de conceito (monitorização da segurança), identificação de dose (menor dose eficaz

e maior dose segura) e farmacocinética em presença de doenças, e; fase tardia ou IIB, realizada com grupos maiores em paralelo com um grupo controle (placebo ou controle ativo), para confirmar os achados dos estudos de fase I e IIA). A fase II é importante para determinar as doses da fase III.

3. Fase III: são estudos de eficácia e segurança com uso de dose definida, por isso chamados de estudos confirmatórios. Apresentam duração de 12-48 meses e envolvem mais de 1000 participantes e, para isso, sendo muitas vezes multicêntricos. O objetivo principal é a validação e confirmação dos resultados de fase II, ou seja, avaliação da eficácia e da segurança à longo prazo, possibilitando sua utilização como evidência científica para registro do medicamento (geração de bulas). Estes estudos avaliam eventos adversos e interações medicamentosas, devendo ser comparados com o tratamento reconhecido como padrão ou, caso não haja, com placebo.

4. Fase IV: corresponde ao estudo após aprovação de registro por órgãos regulatórios como a ANVISA. São estudos realizados pós-comercialização, tendo sua condução a longo prazo e avaliam reações adversas graves. Fornece dados adicionais de segurança, comparação com tratamentos disponíveis e outras formas terapêuticas.

Ensaio clínico randomizado

Os ensaios clínicos randomizados (ECRs) são projetados para responder a uma questão de pesquisa específica e comparar diferentes tratamentos ou intervenções terapêuticas. Há uma população-alvo, da qual será retirada a amostra, seguindo os critérios de elegibilidade definidos pelo pesquisador, enquanto outros indivíduos são excluídos por não satisfazerem os critérios de inclusão.

Esse tipo de delineamento prospectivo conta com dois ou mais grupos, também chamados de "braços de estudo", em que os participantes da pesquisa são alocados aleatoriamente

(randomizados), podendo utilizar-se mais de uma forma de aleatorização, em grupos de intervenção (experimental ou de estudo) e de controle (placebo ou referência), havendo ou não o cegamento (mascaramento) (Figura 1). Ambos os grupos são acompanhados por um período pré-especificado e analisados em termos de desfechos específicos (*end point*), definidos no início do estudo.

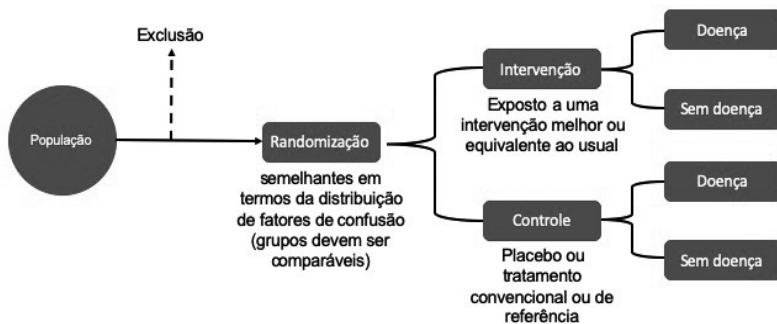


Figura 1. Representação do desenho de um ECR.

Para compreender a relevância metodológica dos ECRs, inicialmente se precisa entender qual é a perspectiva do estudo. Estudos observacionais são dotados de vieses e fatores de confusão por não existir nenhum controle sobre o ambiente estudado. Não é possível obter um entendimento completo da fisiopatologia, dos riscos e benefícios de um tratamento, por exemplo. Nos ECRs, se busca minimizar fatores de confusão e vieses ao realizar a aleatorização da amostra. Nesta perspectiva, tenta-se reproduzir um mundo ideal a partir de um ambiente mais controlado, para garantir que a hipótese do estudo seja testada de uma forma reprodutível. No entanto, pode interferir na validade externa dos resultados visto que os critérios de inclusão e de exclusão são muito restritos nas amostras selecionadas, havendo interferência na generalização dos resultados sobre a população.

Por que o ECR é considerado padrão-ouro?

Os ECRs são, metodologicamente, considerados o padrão-ouro dos estudos. Diferente dos estudos observacionais em que o pesquisador não interfere na exposição, neste delineamento o pesquisador planeja e intervém ativamente nos fatores que influenciam os indivíduos amostrados. São três os aspectos que fazem do ECR o padrão-ouro:

1. é possível avaliar a diferença entre os tratamentos de forma imparcial;
2. a randomização facilita o cegamento do tipo de tratamento alocado e;
3. viabiliza o uso de testes estatísticos (baseados na teoria da probabilidade).

Etapas de um ECR

Existem quatro etapas para a condução de um ECR:

1. fase de seleção: avalia-se se o participante possui as características (critérios de elegibilidade) definidas no protocolo;
2. fase de clareamento (*wash out*) ou de uniformização (*run in*): fase opcional, em que os participantes selecionados recebem ajuste de seus tratamentos de base. Quando os tratamentos de base são diferentes, podem interferir nas avaliações dos resultados introduzindo um erro sistemático, por isso a fase de uniformização;
3. fase de tratamento: os participantes selecionados passam a receber o tratamento ou intervenção, e;
4. fase de acompanhamento (pós-tratamento): garante a segurança do participante no estudo.

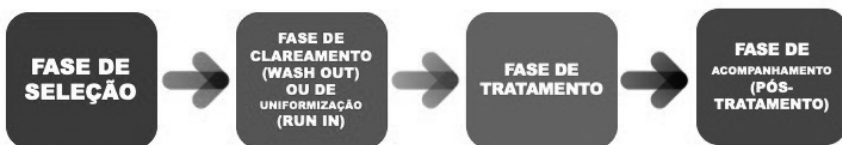


Figura 2. Etapas de um ECR.

Registro do protocolo clínico

Um aspecto extremamente importante na condução de um ensaio clínico é o registro do protocolo clínico. A principal plataforma para efetuar o registro é o ClinicalTrials (clinicaltrials.gov/). Existe também o Registro Brasileiro de Ensaio Clínicos, o ReBEC (www.ensaiosclinicos.gov.br/). O ReBEC é uma plataforma virtual de acesso livre para registro de estudos experimentais e não-experimentais realizados em seres humanos e conduzidos em território brasileiro, por pesquisadores brasileiros e estrangeiros. Todos os detalhes dos estudos registrados no ReBEC estão disponíveis ao público.

Nessas plataformas, além do registro, os pesquisadores podem buscar ECRs similares, evitando duplicidades. Lembre-se que o registro do protocolo propicia transparência, o que gera credibilidade para o estudo, e controle do viés de publicação, visto que o trabalho poderá ser encontrado mais facilmente por grupos que desenvolvem revisões sistemáticas, metanálises e diretrizes clínicas.

Tamanho de amostra de um ECR

O tamanho da amostra é importante para a credibilidade dos resultados fornecidos pelo ECR. Por isso, não deve ser determinado com base em conveniência, arbitrariedade, ou pelo número de indivíduos facilmente recrutados, mas pela necessidade de conferir uma capacidade estatística de detectar uma diferença clinicamente relevante (significativa) entre os grupos.

Uma amostra muito pequena pode impossibilitar a demonstração de uma diferença de efeito entre os grupos, havendo o risco de se chegar a uma conclusão que não corresponde à realidade. Por outro lado, se for utilizado um número superior de sujeitos do que o necessário, algumas pessoas podem estar sendo expostas desnecessariamente a um tratamento menos eficaz ou ao placebo ou a muitos efeitos adversos.

Com a elaboração de um ECR, pretende-se obter um número suficientemente grande de participantes para que um efeito clínico importante seja estatisticamente significativo, mas

não tão grande para que desfechos pouco importantes sejam significativos. Além disso, quando excessivamente grandes, os ECRs podem ser extremamente caros.

Nos casos em que o tamanho de amostra necessário é considerado muito grande, a participação colaborativa é uma alternativa efetiva, realizando o ensaio em vários centros, constituindo o chamado ensaio multicêntrico. Para a condução de um estudo multicêntrico é de vital importância o seguimento do estudo através de um protocolo padronizado.

Critérios de elegibilidade

Assim como outros aspectos que veremos a seguir, a definição clara e detalhada da amostra é de extrema relevância para a boa condução do estudo, pois está diretamente relacionada à generalização dos dados. Os critérios de elegibilidade correspondem aos critérios de inclusão e de exclusão dos participantes da pesquisa. A escolha da amostra deve considerar o objetivo do estudo e ser apoiada na literatura científica.

No ECR, os indivíduos são alocados aleatoriamente (randomizados) em, pelo menos, dois grupos: intervenção e controle. A randomização da amostra ao acaso busca distribuir de forma homogênea os fatores de confusão, tornando os grupos semelhantes. Ou seja, os indivíduos selecionados devem ser comparáveis (similares) em termos da distribuição de fatores de confusão. Essa aleatoriedade não garante a semelhança entre os indivíduos, mas, quanto maior o número de sujeitos em cada grupo, menor a probabilidade desses fatores de confusão ocorrerem.

Braços do estudo

Os grupos de estudo podem ser denominados de “braços do estudo”: controle e intervenção. Em um ECR, o tratamento ou procedimento de interesse pode ser comparado tanto com nenhum tratamento, como a uma terapêutica semelhante (ou seja, de mesma classe, o que chamamos de “*head-to-head trials*”), ou ainda a norma preexistente (tratamento padrão) ou placebo.

Os ECRs são conduzidos sob condições controladas para garantir que a hipótese do estudo seja testada de forma reprodutível. Protocolos mais restritos nas características da amostra podem comprometer a validade externa dos resultados do estudo, pois muitas vezes não podem ser repetidos na prática clínica, ou seja, não podem ser generalizados para outras populações.

Randomização ou aleatorização

No momento da alocação dos participantes, deve-se utilizar um método que torne a distribuição mais homogênea possível por meio de técnicas de aleatorização ou randomização. Técnicas não aleatorizadas não são indicadas, pois pressupõe distorções nos resultados em consequência da casuística ou amostragem, podendo estar viciada por possíveis tendenciosidades. A randomização é um método para selecionar os indivíduos elegíveis para o estudo e dividi-los em grupos, de tal forma que os grupos não sejam afetados pelas características dos participantes, como idade e sexo.

O ECR pode ser proposto nas seguintes condições: a) quando há discordância sobre o tratamento de escolha para o paciente; b) quando não há subsídios para julgar qual das alternativas é preferível no caso específico e; c) o paciente não tem preferência por uma das alternativas. Nestas condições, o paciente é destinado, por acaso, a um tipo de tratamento e, após consentimento informado, passa a ser tratado pelo tipo de tratamento ao qual foi randomizado. Caso um tratamento seja superior ao outro, o paciente não deve ser incluído em estudo randomizado.

Nas técnicas de randomização, após definição dos critérios de inclusão e exclusão, os participantes são sorteados (alocação aleatória) para pertencerem a um dos grupos (controle ou experimental) sem propiciar nenhuma possibilidade de previsão da alocação, ou seja, requer que os participantes tenham igual chance de serem alocados em qualquer um dos grupos. O objetivo da aleatorização consiste em evitar vieses e aumentar a validade dos ensaios clínicos.

Quando implementada adequadamente, a randomização evita o viés de seleção e produz grupos no estudo comparáveis quanto a fatores de risco basais conhecidos e desconhecidos. Os

investigadores e os participantes devem ser incapazes de prever em qual grupo cada um dos participantes será alocado (sigilo de alocação). Além disso, os investigadores devem ser incapazes de alterar a alocação de qualquer participante após a randomização. Existem diferentes formas de randomização e um mesmo estudo pode utilizar mais de uma técnica:

1. alocação simples: é o equivalente ao lançamento de uma moeda: um participante tem igual chance de ser alocado para o grupo de intervenção ou controle. Ao invés do lançamento de uma moeda, entretanto, utiliza-se uma lista de randomização, gerada por computador e utilizada para preparar envelopes selados e sequencialmente numerados. Este tipo de alocação é de baixo custo e de fácil implementação. Por outro lado, pode gerar desequilíbrios no número de participantes nos grupos, bem como na distribuição de fatores de risco na linha de base, principalmente se amostras pequenas;
2. alocação estratificada: é uma alternativa quando se deseja equilíbrio entre os fatores de risco basais, quando há interesse ou necessidade de se estudar, separadamente, sexos, determinados grupos etários ou variantes clínicas e estádios clínicos diferentes da doença. Cada participante é primeiramente classificado em estratos de acordo com características basais pré-definidas (por exemplo, idade, sexo ou gravidade da doença) e cada estrato tem uma lista separada de randomização. Em seguida, os participantes são aleatorizados para grupo de intervenção ou controle. Estratégias de randomização estratificada e em bloco podem ser combinadas: os participantes são categorizados em um estrato, e então, aleatorizados em blocos;
3. alocação em blocos: a lista de randomização é uma sequência aleatória de blocos de participantes em vez de participantes individuais. Os blocos têm um tamanho pré-determinado (blocos fixos e pré-definidos) para a distribuição dos participantes. Utilizado em estudos com baixo recurso e garante que grupo intervenção e grupo controle sejam equilibrados quanto ao número de participantes, mas pode ocorrer prevalência de determinado fator em um dos grupos;

4. alocação ponderal: um número maior de participantes é incluído em um determinado grupo (por exemplo, 2:1 ou 3:1). Esse tipo de alocação é utilizado quando já se tem conhecimento dos benefícios da intervenção e, por questões éticas, um maior número de pessoas se concentra no grupo de intervenção a fim de se beneficiar o maior número de indivíduos;

5. alocação interativa (ou adaptativa): utiliza algoritmos de computador que levam em consideração fatores de risco basais (variáveis de confusão que podem interferir nos resultados) e a alocação dos participantes anteriores para alocar o próximo participante. É utilizado em estudos de desenvolvimento de medicamentos. A vantagem desse método é que ele acomoda mais fatores de risco basais do que a estratificação e, ao mesmo tempo, otimiza o equilíbrio dos grupos. Entretanto, é mais complexa e requer suporte de softwares.

Como escolher o melhor modelo de aleatorização?

Algumas dicas para escolher o melhor método de aleatorização ou considerar a combinação de mais de um:

1. a randomização simples é de fácil implementação e de baixo custo, podendo ser uma boa opção para grandes ensaios ($n > 200$);
2. a estratificação é uma boa opção para proporcionar equilíbrio de covariáveis importantes;
3. já a randomização em bloco é uma alternativa quando se deseja equilíbrio no número de participantes em cada grupo;
4. os métodos de randomização interativa podem ser interessantes quando a estrutura do ensaio inclui estatísticos e suporte de tecnologia da informação.

Para todos os métodos, a implementação adequada é fundamental para garantir o sigilo de alocação e para evitar manipulações e viés de seleção.

Métodos não adequados de randomização incluem, por exemplo, fazer cirurgia por laparoscopia às segundas e quartas-feiras e por laparotomia às terças e quintas-feiras, ou então fazer uma cirurgia por laparoscopia e outra por laparotomia, alternando as vias. Ou também, pelo último dígito da data de nascimento (números pares para o grupo A, ímpares para o grupo B), o lançamento de uma moeda (cara para o grupo A, coroa para o grupo B), a alocação sequencial (paciente A para o grupo A, paciente B para o grupo B, etc.) e a data da consulta (todos os pacientes vistos nesta semana para o grupo A, todos os que consultarem na próxima semana no grupo B, etc.). Esses métodos propiciam falhas na distribuição aleatória, mesmo que inconscientemente.

A randomização bem-feita não garante a homogeneidade dos grupos, principalmente com amostras pequenas. É recomendado verificar, antes da análise dos dados, se houve realmente uma distribuição semelhante dos participantes, particularmente das características que influenciam diretamente no resultado, como estágio da doença, idade, massa corporal, etc. Grupos não homogêneos para os fatores de confusão podem enviesar os resultados.

Placebo e o efeito placebo

O placebo é uma substância ou processo percebido como terapia, mas que não tem atividade biológica ou terapêutica para a doença em estudo. Um placebo pode ser um comprimido que seja idêntico à medicação, mas sem a presença do princípio ativo, ou pode ser um procedimento falso ("*sham procedure*"), como a criação de uma incisão na pele sob sedação para simular um procedimento cirúrgico.

O efeito placebo é uma resposta aprendida em que diferentes tipos de pistas – verbal, condicionada e social – podem gerar expectativas que levam a alterações do comportamento e do resultado clínico via sistema nervoso central, por ações psicológicas e neurobiológicas originadas pelo placebo.

Implicações éticas do uso de placebo

O uso de placebo nem sempre é ético, principalmente nos casos em que coloca em risco o paciente. Quando já há uma terapia padrão efetiva e aceitável, não se deve utilizar placebo. Há exceção para situações triviais para se avaliar o novo tratamento, ou quando o tratamento convencional é fraco e apresenta muitos efeitos adversos.

O uso de placebo é aceitável em estudos onde não existe intervenção atualmente comprovada. Ou, nos casos em que por razões metodológicas convincentes e sólidas, o uso de placebo se faz necessário para determinar a eficácia e segurança de uma intervenção. Nestes casos, quando os pacientes receberem placebo, eles não estão sujeitos a qualquer risco de dano grave ou irreversível.

O placebo não equivale ao não tratamento. Em situações de ausência de terapia padrão, é ético e metodologicamente recomendável a comparação da nova terapia tanto com o placebo, como com o verdadeiro controle. Entretanto, na presença de um tratamento padrão "A", eticamente essa terapia deve ser administrada a todos os participantes. Além do padrão "A", usa-se a nova terapia "B", combinando "A + B" em um grupo e "A + placebo" em outro grupo.

Estudos de não-inferioridade, de equivalência e de superioridade

Estudos de não-inferioridade têm o objetivo de mostrar que um novo tratamento ou procedimento não é menos eficaz que outro já estabelecido e considerado como controle. São de grande importância no estudo de tratamentos em que o uso de placebo é inviável. Estudos de equivalência são destinados a mostrar a ausência de diferenças importantes entre tratamentos (padrão e o que se está testando) e tornaram-se essenciais para a aprovação regulatória dos medicamentos genéricos. Estudos de superioridade têm o objetivo de mostrar que um novo tratamento ou procedimento é mais eficaz que outro já estabelecido.

As diferenças entre os estudos de superioridade, equivalência e não-inferioridade estão, essencialmente, na formulação das hipóteses, nula e alternativa, a serem testadas (Tabela 1).

Tabela 1. Formulação de hipóteses para estudos de superioridade, não-inferioridade e equivalência. Fonte: Pereira Pinto, 2010.

Tipo de estudo	Hipótese nula	Hipótese alternativa
Superioridade	$H_0: [C - T] \geq 0$	$H_A: [C - T] < 0$
Não-inferioridade	$H_0: [C - T] \geq M$	$H_A: [C - T] < M$
Equivalência	$H_0: [C - T] \geq M$	$H_A: [C - T] < M$

H_0 , hipótese nula; H_A , hipótese alternativa; T, medida de eficácia do novo tratamento; C, medida de eficácia do tratamento controle; M, margem de não-inferioridade/equivalência.

Mascaramento ou cegamento

O mascaramento, também conhecido como cegamento, pode ser aplicado aos participantes, aos pesquisadores e aos avaliadores dos resultados do estudo. Em um estudo randomizado, o cegamento é tão importante quanto a randomização, pois protege o estudo de diferenças resultantes de aspectos associados ao tratamento, mas que não fazem parte da intervenção.

O cegamento previne certos vieses que podem afetar a alocação, a avaliação ou a adesão. O conhecimento do participante sobre a designação do tratamento pode afetar a adesão ao regime terapêutico e enviesar percepções de sintomas, comprometendo a avaliação do desfecho. A avaliação de um observador pode ser afetada por saber qual tratamento um participante está recebendo.

Existem diferentes formas de mascaramento:

1. aberto: não ocorre mascaramento, ou seja, tanto o participante como a equipe têm conhecimento do grupo em que o participante foi alocado. É aceito quando o cegamento não é possível, por questões éticas e metodológicas;
2. unicego ou simples cego: o participante não tem conhecimento do tratamento, apenas a equipe de investigação sabe qual foi o tipo de tratamento instituído em cada paciente ou a que grupo cada paciente pertence. É utilizado na impossibilidade de cegamento duplo;

3. duplo-cego: a equipe e os participantes desconhecem a que grupos pertencem, se ao grupo controle ou de intervenção. Apenas uma pessoa da equipe, que não interage com os participantes e tampouco avalia os resultados, sabe quem pertence a cada grupo. É o delineamento mais preciso e confiável, já que é menos vulnerável aos erros de indução. Os grupos de intervenção e controle devem ter a mesma aparência física, como cápsulas com mesma cor e tamanho. No caso do uso de equipamentos, o efeito placebo consistiria, por exemplo, em aplicá-lo desligado ou com uma potência mínima e comprovadamente ineficaz;
4. duplo-placebo: os participantes dos dois grupos recebem as medicações na mesma quantidade e nas mesmas apresentações. O mascaramento permanece íntegro e os dados obtidos precisos. Por exemplo, o grupo "A" recebe o inalatório ativo e o comprimido placebo, e o grupo "B" recebe o inalatório placebo e o comprimido ativo;
5. triplo-cego: além da equipe e dos participantes, o profissional de estatística, que faz as análises, também não sabe qual é o grupo controle e o de intervenção;
6. quadruplo-cego: além da equipe, dos participantes e do estatístico, o investigador que está escrevendo a discussão sobre os resultados, inicialmente de forma genérica quanto à identidade dos grupos, também não sabe qual é o grupo controle e o de intervenção. Somente depois de escrito o artigo, as identidades dos grupos são reveladas, e o artigo será editado para sua versão final.

Implicações éticas do cegamento

Como aplicar um tratamento menos eficaz que outro e garantir o respeito do direito que o paciente tem em saber qual o melhor tratamento para a sua doença ou condição de saúde? Existem importantes implicações éticas no cegamento. No TCLE, o participante deve ser informado sobre todos os grupos de tratamento, inclusive o placebo. Além disso, existem normas e regras para quebra do mascaramento nos casos em que a segurança do paciente estiver comprometida.

Desfechos primários e substitutos

Os desfechos primários, quanti ou qualitativos, são definidos de antemão no protocolo de pesquisa e representam um fator que permite identificar as diferenças entre os grupos, quando comparados, para responder à pergunta de pesquisa. O melhor desfecho é aquele que, de forma satisfatória, mede o evento principal do estudo.

Porém, em alguns casos, pode-se utilizar desfechos substitutos, que correspondem a variáveis mensuradas de forma relativamente fácil e que preveem indiretamente um desfecho raro ou distante de um estímulo tóxico (por exemplo, um poluente) ou uma intervenção terapêutica (por exemplo, um medicamento, procedimento cirúrgico, aconselhamento), mas que não são medidas diretas de dano ou de benefício clínico. Ou seja, correspondem a fatores que buscam prever o evento clínico.

Os desfechos substitutos reduzem, consideravelmente, o tamanho da amostra, a duração do estudo e, conseqüentemente, os custos. Além disso, permitem que os tratamentos sejam avaliados em situações em que o uso de desfechos primários seria, excessivamente, invasivo ou antiético. No entanto, ao se fazer uso de desfechos substitutos, deve-se ficar atento pois podem não responder à questão de pesquisa ou não refletir bem o alvo do tratamento, não sendo válidos ou confiáveis. Têm as mesmas limitações que o uso de qualquer outra medida isolada de sucesso ou falha do tratamento, que é ignorar todas as outras medidas.

Ao optar por um desfecho substituto, deve-se certificar de que este é confiável, reproduzível, clinicamente disponível, facilmente quantificável, economicamente acessível e que exibe um efeito de "dose-resposta", ou seja, quanto maior o nível do desfecho substituto, maior a probabilidade de existência de doença. Assim, este desfecho deve ser um preditor verdadeiro da doença ou do risco de doença e não meramente expressar a exposição a uma covariável.

Adicionalmente, o desfecho substituto deve ser sensível, específico e apresentar um ponto de corte preciso entre os valores normais e anormais e, por fim, deve ser suscetível ao monitoramento para controle de qualidade.

Delineamentos de ECR

Ensaio clínico controlado aleatorizado ou paralelo

É o modelo de ensaio clínico mais utilizado, simples e comum. Os participantes são alocados em um determinado grupo, recebem tratamentos distintos, sendo ambos os grupos incluídos ao mesmo tempo, e seguem nesse mesmo grupo até o encerramento do estudo. São comparados, simultaneamente, os grupos de indivíduos, um dos quais recebe a intervenção e o outro é um grupo controle. Um estudo longitudinal e prospectivo, controlado, aleatorizado e com mascaramento, é considerado a ferramenta mais poderosa para evidenciar a eficácia da intervenção.

Ensaio clínico controlado pareado

Semelhante ao paralelo, pois os tratamentos são estudados simultaneamente, porém há o pareamento dos participantes da pesquisa, combinados em pares por características que podem influenciar no desfecho, como idade e sexo, assegurando que os indivíduos tratados com cada tratamento tenham características semelhantes. Também ocorre em situações em que os tratamentos são investigados simultaneamente no mesmo paciente, ou seja, o sujeito é pareado com ele mesmo. Nestes casos, o próprio participante é o controle de si mesmo. São utilizados quando os tratamentos podem ser dados de forma independente às partes do corpo como, por exemplo, olhos, orelhas, pernas, braços etc.

Ensaio clínico controlado sequencial

Semelhante ao paralelo, porém os grupos são acompanhados até que haja um grupo com claros benefícios ou ambos não apresentem diferença. Utiliza-se esse delineamento quando os resultados precisam ser obtidos com bastante rapidez. Se existir uma diferença grande entre os tratamentos, então esse estudo será realizado em menor tempo quando comparado ao ensaio paralelo. A principal vantagem é que se um tratamento mostrar uma superioridade sobre o outro, o estudo pode ser interrompido.

Ensaio clínico controlado cruzado

Nesse modelo, cada participante recebe tanto o tratamento de intervenção como o de controle, em ordem aleatória,

frequentemente separados por um período sem tratamento (*washout period*). Após a pausa temporal (clareamento), faz-se uma inversão, ou seja, o grupo tratamento passa a receber o controle e, o grupo controle o tratamento, cada grupo sendo o controle de si mesmo.

Ensaio clínico controlado fatorial

Esse modelo corresponde a uma variação do ensaio clínico controlado aleatorizado, porém com um delineamento fatorial, apropriado quando o interesse é investigar a interação entre os fatores. Ao invés de se considerar apenas um fator, (por exemplo, de uma droga ou intervenção "A"), pode-se testar os efeitos da droga ou intervenção "A", droga ou intervenção "B", drogas ou intervenções "A + B", além do placebo ou procedimento controle, formando-se assim, por exemplo, quatro grupos.

Quando o ECR deve ser considerado?

O ECR deve ser considerado quando há incerteza sobre o efeito de uma exposição ou tratamento ou ainda quando a exposição pode ser modificada em um estudo, sem trazer riscos ao participante da pesquisa.

Algumas questões devem ser consideradas antes de se colocar em prática um ECR, a fim de garantir que o estudo seja factível, relevante e ético: o ECR oferece chances de vir a ser executado inadequadamente por ser de difícil e complicada realização? Os objetivos do ECR condizem com um delineamento de ensaio clínico? Há razões para se conduzir o ECR ou existe disponível e é aceito um tratamento ou uma intervenção conhecida como eficaz e superior? Argumentos éticos opõem-se a sugerir randomização? O número de participantes necessário para o estudo é muito grande?

Quando o ECR não deve ser realizado?

Um ECR é considerado desnecessário quando já se tem conhecimento de uma intervenção claramente bem-sucedida para a doença ou condição de saúde de interesse ou quando um ECR ou revisão sistemática com metanálises anteriores produziram resultados definitivos.

São considerados impraticáveis os ECR em que é antiético solicitar consentimento para randomizar ou quando o número de participantes necessários para demonstrar uma diferença significativa entre os grupos for muito grande.

Por fim, um ECR é considerado inadequado nos casos em que os estudos buscam examinar o prognóstico de uma doença (nestes casos, deve-se optar por um estudo longitudinal) e nos estudos que avaliam a validade de um teste diagnóstico ou de rastreamento (nestes casos, o estudo transversal seria o mais adequado).

Vantagens e desvantagens do ECR

Dentre as vantagens, o ECR possibilita a avaliação rigorosa de uma variável única em um grupo precisamente definido de pacientes; um delineamento prospectivo; a alocação aleatória neutraliza prováveis vieses e fatores de confusão; a obtenção de informações de riscos e benefícios e; a avaliação da ocorrência de eventos adversos.

Quanto às desvantagens, tem limitada generalização (validade externa) da população e ambiente em estudo, restringindo a aplicabilidade clínica; são extremamente laboriosos, caros e despendem muito tempo; os participantes podem ter uma linha de base de risco para o desfecho de interesse muito maior do que a média da população ou podem ser oriundos de populações muito diferentes; podem existir falhas em randomizar todos os pacientes elegíveis ou de mascarar os pesquisadores; os desfechos substitutos podem não refletir os desfechos que são importantes para os pacientes; a aplicação em algumas áreas é limitada, como a cirúrgica, na qual não se consegue cegar o cirurgião e, às vezes, dependendo do procedimento, também não é possível cegar o paciente; pode-se apresentar um número insuficiente de indivíduos em subgrupos ou por um tempo demasiadamente curto de seguimento; às vezes, um evento adverso acaba interrompendo o ensaio precocemente e; existe viés de publicação, uma vez que os ECRs com resultados desfavoráveis tendem a não ser publicados.

Quando um ensaio clínico não é randomizado

Alguns artigos que descrevem ensaios clínicos comparando um grupo de intervenção com um grupo controle não são necessariamente randomizados. A terminologia correta é outros ensaios clínicos controlados ou ensaio clínico não randomizado ou ensaio quase-experimental ou quase-randomizado, expressões essas utilizadas para descrever estudos comparativos em que os participantes foram alocados para os grupos de intervenção ou controle de modo não aleatório (não randomizado). Estes estudos são menos efetivos para controlar a influência das variáveis de confusão, as quais nem sempre se consegue ajustar ou eliminar com métodos estatísticos, especialmente, os não mensuráveis.

Nos estudos “pseudorandomizados”, o investigador pode adulterar a pesquisa, através da manipulação da sequência de alocação ou a elegibilidade de novos assuntos. Esta estratégia é utilizada quando a randomização não é ética, nem prática ou for impossível. Se for possível conduzir um ensaio clínico bem controlado e que forneça resultados cientificamente aceitáveis, pode-se não empregar a randomização. De qualquer modo, se o artigo que você estiver lendo é um ensaio clínico não randomizado, você deve usar seu bom senso para decidir se as diferenças na linha de base entre os grupos de intervenção e controle são tão grandes a ponto de invalidar quaisquer diferenças atribuíveis aos efeitos da intervenção.

Opções relacionadas ao “braço controle”, conhecidas como “controle histórico” ou “controle não concorrente”, são utilizados controles de outro momento histórico ou outro lugar geográfico, por exemplo, quando comparam novos pacientes (recebendo o novo tratamento) com pacientes que já foram tratados com outro tipo de tratamento (grupo controle antigo). Esse tipo de estudo também se classifica como ensaio clínico não randomizado. Além disso, um controle histórico pode diferir dos atuais pacientes de várias maneiras e, conseqüentemente, não fornece uma comparação imparcial. O ideal é contar com um braço controle concorrente, diminuindo-se, assim, o risco de vieses.

Alternativas aos ECRs clássicos com dois grupos paralelos, que podem ser utilizados em circunstâncias especiais, incluem:

ensaio fatorial, em que se utiliza mais de dois braços, nos quais se realizam a intervenção em momentos diferentes. Este tipo de pesquisa tenta responder duas ou mais perguntas de pesquisa, na mesma coorte de pacientes e; a randomização de grupo ou *cluster*, na qual a alocação não se realiza de maneira individual, mas, por grupo. Por exemplo, grupo de jogadores de algum esporte, o que permite realizar a intervenção em grupos maiores, diminuindo os custos e permitindo a avaliação do efeito em massa.

Como redigir um ECR?

Com a intenção de demonstrar que os resultados de um ensaio clínico são credíveis, existe uma maneira padronizada de reportar este tipo de estudo. O guia de redação CONSORT (*CONsolidated Standards Of Reporting Trials*), criado em 2010 por uma equipe multidisciplinar, fornece os componentes base para redação de um artigo reportando um ensaio clínico e se encontra disponível na plataforma Equator Network (www.equator-network.org) e no site da CONSORT (www.consort-statement.org). Além disso, existem outros guias personalizados da CONSORT, específicos para ensaios clínicos, como CONSORT Abstracts (para resumos de ECR para periódicos e congressos), CONSORT Non-inferiority (para ECR de não-inferioridade e equivalência), além de outros como o TREND (*Transparent Reporting of Nonrandomized Designs*, para ensaio clínico não randomizado), TIDieR (*Template for intervention description and replication*, um guia sobre como descrever as intervenções de um ensaio clínico), STRICTA (*Standards for Reporting Interventions in Clinical Trials of Acupuncture*, ensaios clínicos de intervenções de acupuntura).

Como avaliar a qualidade metodológica de um ECR?

A qualidade de um ECR depende da adesão do protocolo, tanto dos participantes da pesquisa como dos pesquisadores. Para a avaliação do risco de viés de ECR sugere-se o uso da ferramenta da Colaboração Cochrane (methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials) que revisa, sistematicamente, componentes de riscos do estudo e se há a possibilidade de viés.

Caso seja de interesse compreender a avaliação do risco

de viés de um ECR pela ferramenta, é indicada a leitura do artigo de Viega de Carvalho, Silva & Grande (2013), o qual detalha minuciosamente as características para avaliação de risco de viés em diferentes domínios.

Avaliação crítica de um ECR

Na avaliação crítica de um ECR, devemos considerar pelo menos três fatores: 1) o padrão de recrutamento e seleção; 2) os critérios de inclusão e exclusão e; 3) a generalização das conclusões. Além disso, existem algumas questões norteadoras para a avaliação crítica do ECR, que devem ser consideradas:

1. o estudo foi registrado? O registro do ECR confere: transparência dos dados, fortalecendo o valor ético e científico dos estudos; credibilidade, pois verifica se o estudo foi conduzido em consonância com o protocolo original; obrigação ética com o participante da pesquisa e com a sociedade. O controle de viés traz dados confiáveis e sem ônus para revisões/meta-análises e diretrizes clínicas e; evita duplicidade de esforços e exposições desnecessárias. É muito importante que os protocolos dos ECR sejam registrados e seguidos. O artigo deve informar o número de registro do protocolo, que pode ser feito em plataformas como ClinicalTrials e ReBEC;

2. os participantes selecionados são semelhantes aos pacientes na vida real? É possível a generalização dos resultados? O ECR deve descrever claramente quais foram os critérios de inclusão e exclusão. Critérios rígidos de inclusão são mais restritivos na generalização das conclusões, entretanto são mais específicos. É importante lembrar que apenas será possível generalizar para as populações semelhantes à do estudo em questão. Caberá ao leitor decidir se as diferenças entre a população selecionada no estudo e o seu público de interesse são importantes. Deve-se mencionar como os participantes foram selecionados da população geral;

3. quais medidas de desfecho foram usadas? O desfecho foi bem descrito? O desfecho é adequado para a avaliação de interesse? Os eventos clinicamente relevantes foram considerados? Embora muitos trabalhos sejam

metodologicamente bem desenvolvidos, estudam variáveis não muito úteis clinicamente ou que não representam aquilo que deveriam. O tratamento deve ter impacto sobre os desfechos, sobre os eventos clínicos de maior interesse para o paciente e para o seu médico. Relevância clínica vai além da estatística e é determinada por julgamento clínico baseado em evidências estatísticas e por isso é necessário a definição clara de um desfecho relevante. Muitos estudos utilizam desfechos substitutos, que constituem uma medida indireta em substituição ao desfecho clínico. A vantagem de se utilizar o desfecho substituto é que o tamanho da amostra pode ser reduzido, assim como o custo e a duração do estudo. As desvantagens são que, muitas vezes, estas medidas escondem efeitos deletérios do tratamento em outros desfechos, talvez clinicamente mais relevantes, que podem se manifestar apenas em maior tempo de seguimento;

4. como foi feita a distribuição dos indivíduos entre o grupo tratamento e o grupo controle (ou placebo)? Os grupos eram semelhantes, exceto pela intervenção? Trabalhos que não incluem grupo controle ou placebo perdem credibilidade, pois correm o risco de mostrar, equivocadamente, que um tratamento é eficaz. Ainda, a utilização do placebo pode estar eticamente inadequada em alguns estudos. Uma alternativa para evitar esse tipo de erro é comparar o grupo com o novo medicamento com outro com o tratamento considerado padrão. Em situações que o uso do placebo é inadequado, também convém proceder com estudos de não-inferioridade. De uma maneira geral, as técnicas de aleatorização tendem a evitar que os grupos de pacientes sejam heterogêneos. Entretanto, quando um estudo possui um tamanho de amostra inadequado, essa homogeneidade não estará garantida. Assim, quanto maior o tamanho da amostra, maior é a garantia que características e fatores relacionados ao desfecho do estudo tendam a ser igualmente distribuídos entre os grupos;

5. como foi feita a randomização entre os grupos tratamento e controle? A randomização implica a distribuição aleatória dos pacientes, de modo que cada indivíduo possa ter a mesma chance de ser alocado nos diferentes grupos.

Estudos controlados não randomizados perdem qualidade na evidência científica, pois, na maioria das vezes, os grupos não são homogêneos;

6. a intervenção avaliada foi, de fato, a única intervenção relevante no paciente? Muitas vezes a utilização de tratamentos concomitantes ao medicamento utilizado pode interferir na eficácia do mesmo, quer por questões farmacodinâmicas, quer por questões de efeitos sinérgicos ou antagônicos à medicação utilizada. Tal conceito é definido como co-intervenção. Uma maneira de tentar controlar tal situação é a inclusão da medicação conhecida interferente nos critérios de exclusão ou a anotação periódica de toda medicação concomitante administrada no paciente ao longo do estudo, para serem consideradas na análise e interpretação dos resultados;

7. a mensuração do evento final de interesse foi realizada sob o mascaramento duplo-cego? O mascaramento é o mais recomendado, pois evita a tendenciosidade dos dados. Mensuração do desfecho pode ser influenciada pelo conhecimento da alocação (viés de observação). Tanto os pacientes quanto os investigadores podem ser afetados, em particular para desfechos subjetivos como dor;

8. o tempo de seguimento (*follow-up*) foi suficiente? O seguimento foi completo? Essa pergunta contribui para compreender se o tempo de estudo foi suficiente para avaliar o evento final de interesse. É igualmente relevante verificar se o percentual de pacientes que abandonaram o seguimento foi semelhante entre os diversos grupos. Algumas fontes potenciais de perda de seguimento devem ser consideradas: alguns pacientes podem não ter a doença como se imaginava no começo do estudo; outros abandonam a pesquisa; não aderem ao tratamento; apresentam reações adversas ou; perde-se o contato. Nestes casos, a comparação entre os grupos de estudo torna-se enviesada, mesmo que o delineamento esteja adequado;

9. o estudo tem cálculo de tamanho de amostra? O tamanho da amostra é adequado para a proposta do estudo? Foi viável? Um ensaio deve recrutar um número

suficiente de pacientes para evidenciar a eficácia da intervenção. Informação importante que pode indicar se houve dificuldades no recrutamento ou falha de protocolo. Para se calcular o tamanho da amostra, deve-se levar em consideração como o desfecho será medido, se quantitativo ou se qualitativo, nível de significância e, poder do teste. Se o desfecho for quantitativo, será necessário ter ideia da magnitude de efeito e desvio padrão. Se o desfecho for qualitativo, será necessário saber qual a proporção de pacientes com desfecho é esperada no grupo da intervenção e qual a proporção de pacientes com desfecho é esperada no grupo controle;

10. os resultados foram analisados com base na intenção de tratar (ITT)? Todos os participantes do estudo, incluindo os que desistiram antes do seu término, devem ser analisados junto com os dados dos pacientes que seguiram todo o protocolo. Deverá ser averiguado se todos os dados sobre os participantes, originalmente alocados ao braço da intervenção do estudo, incluindo os indivíduos que saíram antes do término, os que não tomaram os comprimidos, e mesmo aqueles que, subsequentemente, receberam a intervenção-controle por qualquer razão, devem ser analisados junto com os dados sobre os pacientes que seguiram todo o protocolo. De modo inverso, os participantes, que se retiraram do grupo placebo do estudo, devem ser analisados, conjuntamente, àqueles que tomaram fielmente seu placebo. Este tipo de análise indica fidelidade ao protocolo e é, geralmente, utilizado em estudos de superioridade. O MITT (*modified intention-to-treat*) também pode ser utilizado, este inclui todos os pacientes na análise, exceto aqueles que não receberam nenhuma dose do tratamento. Este tipo de análise é ambígua e pode induzir a interpretação equivocada, além da exclusão inapropriada de participantes, não sendo recomendado pelo CONSORT;

11. como foram apresentados os resultados? Como foi medida a magnitude de efeito da intervenção? Qual é a precisão do efeito do tratamento? Para se avaliar a relevância

clínica, é necessário considerar a estimativa da magnitude do efeito e qual é a precisão da estimativa. As medidas de efeito podem ser medidas por uma razão, sendo denominada de medida relativa, ou por uma diferença, considerada medida absoluta. Para medir a magnitude da associação (força do efeito) entre a intervenção e o desfecho (quantas vezes a ocorrência do desfecho é maior no grupo intervenção em relação ao grupo controle) devemos usar medidas de associação relativas. As medidas de associação por diferença avaliam o quanto a frequência de um desfecho no grupo intervenção excede ao grupo controle, ou seja, avaliam qual a incidência do desfecho atribuído a intervenção. As principais formas de medir o efeito são: risco absoluto (probabilidade de desenvolver o desfecho em cada grupo), redução absoluta de risco (diferença de risco absoluto entre o grupo controle e entre o grupo tratado), risco relativo (risco de eventos entre pacientes no grupo tratado, relativo ao risco nos pacientes do grupo controle), redução de risco relativo (estimativa da proporção do risco basal que é removido pelo tratamento experimental), número necessário para tratar (medida utilizada para avaliar a significância clínica, expressa o número de pacientes que devem ser tratados por um período de tempo para obter um evento favorável ou para prevenir um evento desfavorável). Além disso, o intervalo de confiança de 95%, que nos indica qual o grau de precisão da estimativa do estudo, deverá ser apresentado. Quanto mais estreito, mais precisão o resultado terá;

12. houve diferenças entre os diferentes subgrupos quanto ao desfecho? Foi explicada essa diferença? Caso contrário, pode conferir alteração de tratamento, falha de protocolo, viés ou fator de confundimento não previsto;

13. se os achados foram negativos, o quanto pode ser explicado por falha na implementação e/ou otimização inadequada da intervenção? Um ECR não está livre de limitações metodológicas, importante as identificar;

14. quais pesquisas adicionais os autores acreditam ser necessárias? Se refere ao relato dos autores sobre perspectivas futuras a fim de contribuir com melhorias

metodológicas para as próximas evidências científicas. Os autores podem descrever condutas errôneas que precisam ser consideradas em próximos estudos.

Referências

Bonita, R.; Beaglehole, R.; Kjellström, T. Epidemiologia básica. 2. ed. São Paulo: Santos, 2010-2019. 213 p.

Buchler et al. Como avaliar criticamente um ensaio clínico de alocação aleatória em terapia intensiva. Rev Bras Ter Intensiva. 2009; 21(2):219-225.

ClinicalTrials. Submit Studies: How to Register Your Study. Disponível em: <<https://clinicaltrials.gov/>>. Acesso em: 10 de out. 2020.

CONSORT 2010 Explanation and Elaboration Document. Disponível em: <<http://www.consort-statement.org>>. Acesso em: 10 de out. 2020.

Equator Network. Reporting guidelines for main study types. Disponível em: <www.equator-network.org>. Acesso em: 10 de out. 2020.

Fanaroff et al., Randomized Trials Versus Common Sense and Clinical Observation. Am Coll Cardiol 2020;76:580–9.

Greenhalgh, T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Guimarães, L.S.P.; Hirakata, V.N.; Camey, S.A.; Nunes, L.N.; Mancuso, A.C.B. Os principais delineamentos na epidemiologia. Ensaio clínico (parte II). Rev HCPA 2013;33(3/4):295-302.

Mancuso, A.C.B.; Camey, S.A.; Nunes, L.N.; Hirakata, V.N.; Guimarães, L.S.P. Os principais delineamentos na epidemiologia. Ensaios clínicos (parte I). Rev HCPA 2013;33(3/4):286-294.

Pacheco et al., Guidelines para publicação de estudos científicos. Parte 3: Como publicar ensaios clínicos. Diagn Tratamento. 2017;22(4):169-75.

Parente, R.C.M.; Pinho de Oliveira, M.A.; CELESTE, R.K. Qual é o valor do placebo em pesquisas clínicas? FEMINA. 2011;39(4):177-182.

Pinho de Oliveira, M.A.; Velarde, L.G.C.; Moreira de Sá, R.A.. Ensaios clínicos randomizados: Série Entendendo a Pesquisa Clínica 2. FEMINA. 2015;43(1):7-11.

Pinto, V. F. Estudos clínicos de não-inferioridade: fundamentos e controvérsias. Vasc Bras. 2010;9(3):145-151.

Registro Brasileiro de Ensaios Clínicos. Disponível em: <www.ensaiosclinicos.gov.br/>. Acesso em: 10 de out. 2020.

RoB 2: A revised Cochrane risk-of-bias tool for randomized trials. Disponível em: <<https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials>>. Acesso em: 10 de out. 2020.

Viegas de Carvalho, A.P.; Silva, V.; Grande, A.J. Avaliação do risco de viés de ensaios clínicos randomizados pela ferramenta da colaboração Cochrane. Centro Cochrane do Brasil (CCB). Diagn Tratamento. 2013;18(1):38-44.

Estudos de testes diagnósticos ou de rastreamento

*Johanna Ovalle Diaz
Lilian de Sá Paz Ramos
Edison Capp
Otto Henrique Nienov*

Quando estamos diante de um novo teste diagnóstico ou de rastreamento, é comum questionarmos se: o novo teste diagnóstico ou de rastreamento é bom? Funciona? É mais barato do que o de referência? É realmente necessário? A avaliação da qualidade de testes diagnósticos ou de rastreamento é um tema de interesse da investigação clínica e epidemiológica.

O desempenho de um teste diagnóstico ou de rastreamento depende da ausência de viés e da precisão. Para avaliarmos um novo teste utilizamos parâmetros, como sensibilidade, especificidade e acurácia. No entanto, na prática, a maioria dos testes diagnósticos ou de rastreamento utilizados não é perfeita e produz resultados falso-positivos (o teste é positivo, mas o paciente não tem a doença) e resultados falso-negativos (o teste é negativo, mas o paciente tem a doença). Portanto, aprender a interpretar as propriedades dos testes é essencial.

Testes de rastreamento

Os testes de rastreamento, também chamados de triagem ou de *screening*, não são utilizados para estabelecer um diagnóstico, mas para determinar a presença ou ausência de uma doença ou fator de risco em indivíduos assintomáticos. Estes testes separam os indivíduos aparentemente saudáveis, mas que apresentam um fator de risco para uma doença ou condição de saúde, daquelas que não o apresentam.

A aplicação de testes de rastreamento deve ser norteadada pela premissa de que o diagnóstico precoce, de fato, pode

contribuir para a melhora do prognóstico da doença ou condição de saúde do indivíduo. Desta forma, um bom teste de rastreamento deve ter alta sensibilidade, para não perder os casos da doença presentes na população testada, bem como alta especificidade, a fim de reduzir o número de pessoas com resultados falso-positivos, que necessitam de uma maior investigação posterior.

No entanto, dificilmente encontramos testes que apresentem tanto alta sensibilidade como especificidade. Como o objetivo dos testes de *screening* visa afastar um diagnóstico, desejamos um teste mais sensível do que específico, porque se o resultado for negativo, haverá pouca chance de ser um falso-negativo.

O principal propósito de se desenvolver um novo teste de rastreamento é detectar o maior número de casos com baixo custo, além de não acarretar reações adversas ou efeitos colaterais, devendo este ser mais seguro que os testes diagnósticos. Portanto, um teste ideal para rastreamento deve ser rápido, barato, aceitável pela população, de fácil aplicabilidade e que cause desconfortos e/ou transtornos mínimos aos pacientes e, ainda, que os resultados sejam válidos, confiáveis e reprodutíveis.

São exemplos de testes de rastreamentos o teste do pezinho em recém-nascidos para detectar fenilcetonúria, glicemia de jejum para rastreamento de diabetes, aferição da pressão arterial para hipertensão, mamografia para detecção de câncer de mama e exame de Papanicolau para displasia cervical ou câncer do colo do útero.

Testes diagnósticos

Os testes de diagnóstico são utilizados para determinar a presença ou ausência da doença ou condição de saúde, quando um indivíduo apresenta sinais ou sintomas suspeitos da doença, ou seja, o teste objetiva auxiliar na possível confirmação diagnóstica. Tais testes, também, são utilizados para avaliar a gravidade de um quadro clínico, estimar o prognóstico e monitorar a resposta a uma intervenção.

Os instrumentos para diagnóstico tendem a ser mais precisos e trazer maior detalhamento em relação ao agravo ou

doença ao qual se está investigando. Para isto, se requer um teste mais específico, pois se o resultado for positivo, haverá grande chance de ser um verdadeiro-positivo.

São exemplos de testes diagnósticos o teste oral de tolerância à glicose para diabetes, monitorização ambulatorial da pressão arterial (MAPA) para hipertensão, biópsia percutânea guiada por estereotaxia e ultrassonografia para o câncer de mama e, raspagem endocervical e biópsia para câncer do colo do útero.

As vezes, para se estabelecer um diagnóstico, não há um instrumento único 100% sensível e específico, então torna-se necessário usar testes associados. Assim, por exemplo, para uma mulher com nódulo suspeito na mama pode ser necessária a realização de exames associados para definição do diagnóstico, tais como a mamografia, ultrassonografia e biópsia percutânea.

Resumidamente, o teste diagnóstico é utilizado quando o paciente tem alguns sintomas sugestivos da doença em questão, enquanto que o teste de rastreamento é usado em pessoas sem sintomas que são retiradas de uma população com prevalência muito menor da doença. Desta forma, geralmente, o teste diagnóstico é realizado após um teste de triagem positivo, a fim de estabelecer um diagnóstico definitivo.

Teoria da probabilidade

A teoria da probabilidade (bayesiana) estuda os padrões de fenômenos aleatórios, cuja ocorrência não é conhecida com certeza, tais como eventos aleatórios, variáveis aleatórias e suas propriedades. O estabelecimento de um diagnóstico é um processo imperfeito, ou seja, de probabilidade e não de certeza. Assim, a possibilidade de um paciente ter a doença é expressa pela sua probabilidade de ocorrer. Deste modo, os testes diagnósticos compreendem uma importante aplicação da teoria da probabilidade.

Acurácia, precisão e generalização

Dentre as propriedades que devem ser consideradas na avaliação dos testes diagnósticos e de rastreamento, destacam-se a acurácia (ou validade), precisão e generalização.

A acurácia (ou validade) consiste na capacidade de uma medida ser correta na média. Desta forma, categoriza-se adequadamente os indivíduos doentes como teste positivo e os não doentes como teste negativo. A validade informa se os resultados representam a "verdade" ou o quanto se afastam dela. É um parâmetro que depende da sensibilidade, especificidade e da prevalência da doença. Quando uma medida não é considerada acurada, significa que é viciada.

A precisão, também conhecida como reprodutibilidade, repetibilidade ou confiabilidade, representa a capacidade de a medida dar o mesmo resultado em medições repetidas independentes e sob as mesmas condições. Trata-se da consistência de resultados quando o exame se repete. A comparação dos resultados pode ser apresentada através da taxa global de concordância, entre os examinadores, pelo índice Kappa. A reprodutibilidade dos testes diagnósticos ou de rastreamento também deve ser considerada. Esta propriedade está atrelada à possibilidade de que o teste seja aplicado por profissionais (ou não, de acordo com a sua natureza) e pacientes, em momentos diversos, possibilitando um mesmo resultado.

Também é importante ressaltar que somente podem ser extrapolados (generalizados) dados de estudos para populações semelhantes àquelas pesquisadas e, quando há seleção demasiada dos indivíduos que irão compor a amostra, estes indivíduos podem não representar a população em geral. Tais fatores restringem a generalização, produzindo testes, instrumentos, métodos de diagnóstico ou de rastreamento que demonstram resultados satisfatórios em ambiente de pesquisa, contudo, podem não apresentar respostas semelhantes em outros cenários.

A validade, precisão e generalização são qualidades importantes para os testes de rastreio e diagnóstico. Caso ausentes, os dados tornam-se inúteis, pois não seriam, necessariamente, fidedignos.

Validação de testes diagnósticos ou de rastreamento

A validação consiste em avaliar a utilidade (performance) de um novo teste diagnóstico ou de rastreamento, chamado de

teste índice. O teste índice é aquele que o pesquisador julga ser um instrumento mais barato, conveniente, de fácil execução e interpretação, aceitável para os pacientes e que produz um resultado dicotômico (positivo ou negativo) instantâneo. O teste diagnóstico ideal deveria fornecer, sempre, a resposta correta, ou seja, um resultado positivo nos indivíduos com a doença e um resultado negativo nos indivíduos sem a doença. Para determinar a validade do teste índice, compara-se os resultados deste com os de um teste padrão.

O teste padrão-ouro, também chamado de teste de referência ou *gold standard*, é aquele considerado como o melhor teste disponível, que possui menor probabilidade de erro e deve ser escolhido criteriosamente. Esse teste de referência tem a finalidade de fornecer a medida "verdadeira", ou seja, nos dar a garantia de quem realmente está doente e quem não está. Na prática, nem sempre o teste padrão-ouro pode ser utilizado. Em consequência disso, muitos estudos concentram-se na obtenção de um novo teste.

Os instrumentos são validados de acordo com a população à qual se destinam. São selecionados dois grupos: aqueles com a doença, conforme diagnóstico feito pelo teste padrão, e outro sem a doença, em que o teste de referência foi negativo. O pesquisador que for testar o novo teste deverá estar cegado, ou seja, ao fazer a avaliação do teste índice deve desconhecer os resultados obtidos com o padrão-ouro. O novo teste é, então, realizado em ambos os grupos.

Tabela de contingência 2 X 2

Os resultados do novo teste, em comparação com os do teste padrão-ouro, são apresentados em uma tabela de contingência 2×2 , também conhecida como matriz 2×2 (Figura 1). Desta forma, é possível saber pela alocação dos indivíduos na tabela de contingência, para cada indivíduo, se o resultado do teste de rastreamento combinou com o teste padrão. Este exercício é conhecido como estudo de validação.

		Diagnóstico "verdadeiro" ou teste padrão-ouro		
		DOENTE	NÃO DOENTE	
Teste índice	TESTE POSITIVO	A (verdadeiro-positivo)	B (falso-positivo)	A + B
	TESTE NEGATIVO	C (falso-negativo)	D (verdadeiro-negativo)	C + D
		A + C	B + D	n (A + B + C + D)

Figura 1. Notação da tabela de contingência 2 X 2 para expressar os resultados de um estudo de validação para um novo teste diagnóstico ou de rastreamento.

Na população examinada (Figura 1), os casos situados na coluna esquerda da tabela são positivos para o teste padrão-ouro e, portanto, têm a doença em estudo (verdadeiros doentes); os casos à direita são negativos para o padrão-ouro e, portanto, não têm a doença (verdadeiros não doentes). Por outro lado, os resultados positivos e negativos para o novo teste (teste índice) estão apresentados em disposição horizontal. Dessa maneira, têm-se quatro tipos de resultados: verdadeiros-positivos ("A", indivíduo doente e teste positivo), falsos-positivos ("B", indivíduo não doente e teste positivo, o que representa a possibilidade do erro tipo I), falsos-negativos ("C", indivíduo doente e teste negativo, o que representa a possibilidade do erro tipo II) e verdadeiros-negativos ("D", indivíduo não doente e teste negativo).

A configuração da tabela 2 X 2 fornece, ainda, o total de cada linha e cada coluna, em que temos nas colunas, o número total de doentes ("A + C") e não doentes ("B + D"); nas linhas, o número total de testes positivos ("A + B") e negativos ("C + D") e; na diagonal, o número total da amostra ("n", ou seja, "A + B + C + D").

Sensibilidade e especificidade

Qual é a capacidade do teste novo para detectar pessoas com a doença? Qual é a capacidade do teste novo para excluir corretamente as pessoas sem a doença? As respostas podem

ser obtidas pela sensibilidade ou taxa de verdadeiros-positivos (teste positivo entre os doentes) e pela especificidade ou taxa de verdadeiros-negativos (teste negativo entre os saudáveis).

A sensibilidade e a especificidade são medidas importantes de um teste diagnóstico, porque evidenciam uma ideia do quão eficiente é o desempenho do novo teste, em geral, comparando-se com a performance de um teste padrão-ouro. São utilizados para avaliar a acurácia ou validade do teste. Assim, ambas medidas, geralmente, são usadas em estudos epidemiológicos, de larga escala e com grande tamanho amostral.

A sensibilidade corresponde a proporção entre o número de doentes que o teste consegue detectar ("A") e o número total de doentes ("A + C"). Ou seja, é a probabilidade de o teste ser positivo para uma pessoa doente. Testes sensíveis são úteis para reduzir possibilidades diagnósticas (testes de triagem ou de *screening*), pois apresentam poucos falso-negativos.

A especificidade se refere a proporção entre o número de pessoas não doentes que o teste detecta como "negativas" ("D") e o número total de pessoas não doentes ("B + D"). Ou seja, é a probabilidade de o teste ser negativo para uma pessoa não doente. Testes muito específicos são usados para confirmar a presença da doença, pois apresentam poucos falso-positivos.

É importante alcançarmos um equilíbrio entre a sensibilidade e a especificidade de um teste. Nem sempre será possível obter alta performance destes parâmetros simultaneamente, logo, a razão aceitável entre estas medidas está relacionada ao contexto e objetivo do teste novo proposto.

A escolha do ponto de corte do teste índice depende da importância dada aos falso-positivos e falso-negativos para a doença ou condição de saúde de interesse. Por exemplo, para as triagens sorológicas em bancos de sangue para prevenção de transmissão de infecções, nas quais a não detecção de casos acarretará risco para a população, o ponto de corte deverá ser estabelecido, tendo como objetivo alcançar maior sensibilidade do teste para que não ocorram resultados falso-negativos. Por

outro lado, nas situações de indicações cirúrgicas, deve-se evitar resultados falso-positivos, ou seja, o ponto de corte deve ser definido de tal forma que aumente a especificidade do teste.

Curva ROC (*Relative Operating Characteristics*)

Como determinar o ponto de corte de equilíbrio entre sensibilidade e especificidade? A curva ROC é uma ferramenta poderosa para medir e especificar o desempenho dos testes diagnósticos, por permitir expressar graficamente a relação, normalmente antagônica, entre sensibilidade e especificidade e é utilizada quando se deseja transformar um teste com resultados numéricos contínuos em dicotômicos, através da determinação de um ponto de corte ou *cutoff point*.

A curva ROC demonstra, no eixo Y, a proporção de verdadeiros-positivos (sensibilidade) e, no eixo X, a proporção de falso-positivos (1-especificidade). Desta forma, a curva ROC é um gráfico de sensibilidade ou taxa de verdadeiros-positivos versus a taxa de falsos-positivos (Figura 2).

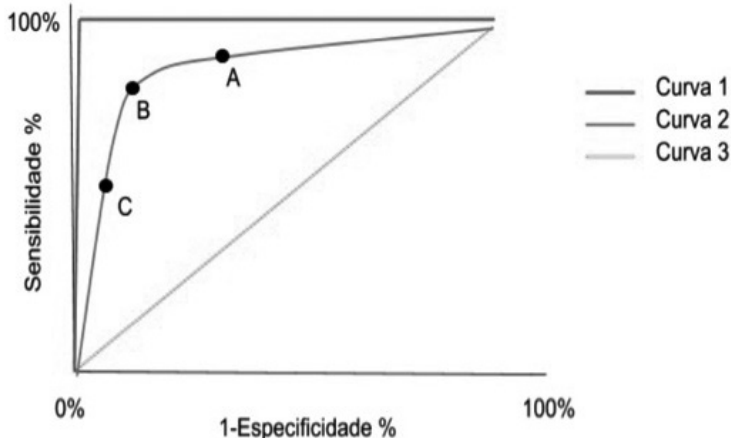


Figura 2. Representação de uma curva ROC.

Esse gráfico possibilita fazer uma avaliação global do valor de um teste diagnóstico por meio do cálculo da área sob a curva (ASC). Os valores da ASC podem variar de 0 a 1, em que valores > 0,8 indicam que a precisão do teste diagnóstico é muito boa.

O valor do ponto de corte é definido com um valor que pode ser selecionado arbitrariamente pelo pesquisador entre os valores possíveis para a variável de decisão, acima do qual o paciente é classificado como positivo (teste positivo) e abaixo do qual é classificado como negativo (teste negativo). Para cada ponto de corte, são calculados valores de sensibilidade e especificidade, que são dispostos no gráfico. O ideal seria dispor de um teste 100% sensível e 100% específico, isto é, que fornecesse resultados sempre positivos, no caso de doença, e sempre negativos, em indivíduos não afetados pela doença, entretanto, isto não ocorre na prática.

O valor escolhido como ponto de corte irá influenciar as características do teste. Conforme demonstrado na Figura 2, observa-se os seguintes tipos de curva ROC: a curva 1 representa o ponto de maior acurácia do teste. Corresponde a um "teste perfeito", no qual, haveria uma separação completa de valores entre pacientes com e sem a doença. O valor de corte seria o menor valor em pacientes com a doença e a ASC corresponderia ao valor 1. No entanto, como não há testes perfeitos, sempre haverá resultados falso-positivos ou falso-negativos. Quanto mais preciso for um teste, maior será a ASC, que é a probabilidade de que o valor da medida seja maior em uma pessoa qualquer com a doença do que em uma pessoa qualquer sem a doença. A curva 3 (linha diagonal do gráfico), expressa a acurácia de 50% e corresponde a um teste que é positivo ou negativo, aleatoriamente (ao acaso).

A curva 2 representa um teste com acurácia de 85%, que significa que em 85% das vezes em que for utilizado, o teste discriminará os verdadeiros-positivos e os verdadeiros-negativos. No entanto, isto também significa que em 15% das vezes fornecerá um falso resultado. Quanto menor o *cutoff point* (ponto "A"), maior é a sensibilidade, porém menor é a especificidade. Neste caso, se pretende obter um teste muito sensível e menos específico, escolhe-se o ponto "A", resultando uma menor proporção de falso-negativos e uma maior proporção de falso-positivos. O ponto de corte localizado no ponto "B" permite pôr em evidência os valores para os quais existe maior otimização da sensibilidade em função da especificidade. Já quanto maior o *cutoff point* (ponto "C"), maior especificidade, porém menor sensibilidade. Neste caso, se o intuito é ter um teste muito

específico e menos sensível, escolhe-se o ponto "C", obtendo-se uma menor proporção de falso-positivos e uma maior proporção de falsos-negativos.

Definido o ponto de corte e avaliada a sensibilidade e especificidade, como decidir qual teste deve ser preferencialmente empregado? Para a escolha do teste, além dos parâmetros de sensibilidade e especificidade, deve-se avaliar acurácia, prevalência, probabilidade pré-teste, valores preditivos (probabilidades pós-teste) e razões de verossimilhança.

Acurácia

Qual foi a proporção de todos os testes que produziu resultados corretos (verdadeiros-positivos e verdadeiros-negativos)? Essa pergunta poder ser respondida pelo cálculo da acurácia, que é a relação entre todos os resultados verdadeiros-positivos e negativos (" $A + D$ ") e o total da população (" n ").

Prevalência

A prevalência indica qual é o percentual de pessoas que realmente estão doentes (" $A + C$ ") na população (" n "). Baixas prevalências tendem a estar associadas a baixos níveis de reprodutibilidade.

Probabilidade pré-teste

A probabilidade pré-teste corresponde a probabilidade de ter uma condição-alvo antes de um teste diagnóstico. Ou seja, é o número total de testes positivos (" $A + B$ ") dividido pela população (" n ").

Valores preditivos positivo e negativo

Se uma pessoa apresentar um resultado positivo, qual é a probabilidade de que ela tenha a doença? E, se uma pessoa apresentar um resultado negativo, qual é a probabilidade de que ela não tenha a doença? As respectivas soluções para estas questões correspondem ao valor preditivo positivo (VPP) e o valor preditivo negativo (VPN), respectivamente. Os valores preditivos,

também denominados de probabilidades pós-teste de um teste positivo e de um teste negativo, nos informam a probabilidade de doença, dado o resultado do teste. Enquanto a sensibilidade e a especificidade de um teste são praticamente constantes, qualquer que seja a prevalência do problema, os VPP e VPN dependem, diretamente, da prevalência.

O cálculo do VPP é obtido através da relação entre a quantidade de pessoas doentes que testaram positivo ("A") e o número total de testes positivos ("A + B"). Ou seja, é a probabilidade de estar doente, dado que o teste deu positivo. Quando um teste diagnóstico tem VPP alto, um paciente cujo teste apresente resultado positivo, muito provavelmente tem a doença que está sendo investigada.

O VPN refere-se à relação entre a quantidade de pessoas não doentes que testaram negativo ("D") e o número total de testes negativos ("C + D"). Ou seja, é a probabilidade de não estar doente, dado que o teste resultou negativo. Quando um teste tem VPN alto, um paciente cujo teste apresente resultado negativo muito provavelmente não tem a doença que está sendo investigada.

Como mencionado, o VPP e o VPN de um novo teste dependem da prevalência da doença na população-alvo, mas também de suas propriedades intrínsecas. Portanto, valores preditivos não são características fixas do teste e não podem ser generalizados para populações com diferentes prevalências da doença. Se a prevalência da doença for alta em uma determinada população, o VPP aumenta e o VPN diminui. Quanto mais sensível, maior o VPN (redução de falsos-negativos, ou seja, maior segurança de que o paciente com teste negativo não tem a doença) e, quanto mais específico, maior o VPP (redução de falsos-positivos, ou seja, maior a segurança de que a pessoa com teste positivo tem a doença).

Razões de verossimilhanças positivo negativo

E, agora, qual é a probabilidade de encontrar um teste positivo em uma pessoa com a doença, comparando com uma pessoa que não o apresenta? Ou, qual é a probabilidade de encontrar um teste negativo em uma pessoa sem a doença, comparando com uma pessoa que o apresenta? Ambas questões

são respondidas pela razão de verossimilhanças de um teste positivo (RV+) e razão de verossimilhança de um teste negativo (RV-), respectivamente.

As razões de verossimilhança combinam sensibilidade e especificidade para quantificar o quão útil um novo teste diagnóstico é para aumentar ou diminuir a probabilidade de ter uma doença em comparação com a prevalência dessa doença (probabilidade pré-teste) na população-alvo. A razão de verossimilhança possui enorme valor prático, sendo útil para incluir ou excluir determinado diagnóstico. As razões de verossimilhança devem ser usadas como chances mais do que como probabilidades.

A RV+ é a probabilidade de um resultado positivo em alguém com a doença dividida pela probabilidade do mesmo resultado em alguém sem a doença. A RV- é a probabilidade de um resultado negativo em alguém com a doença dividida pela probabilidade do mesmo resultado em alguém sem a doença. A RV+ varia de 1 a infinito, e uma RV+ igual a 1 indica que a probabilidade de resultado positivo do teste é a mesma para pacientes com e sem a doença, ou seja, o resultado do teste é inútil. Uma RV+ maior que 1 corrobora com a presença da doença. Quanto maior a RV+, maior será a probabilidade de que o resultado positivo do teste aumente a probabilidade de doença se o resultado do teste for positivo. A RV- varia de 1 a 0, e quanto mais próxima de 0, menor será a probabilidade de doença na presença de resultado negativo do teste.

Qual parâmetro é mais útil para avaliar um teste índice?

A resposta é: depende! Cada parâmetro descreve uma característica específica do teste, e a utilidade de cada medida dependerá de como o teste será utilizado. Se os valores das várias propriedades de um teste, tais como a sensibilidade e a especificidade, se situarem dentro de limites razoáveis, pode-se dizer que o teste foi válido. Um teste totalmente incapaz de discriminar indivíduos doentes e não doentes, teria uma ASC de 0,5, que corresponde a hipótese nula. Por outro lado, se a ASC estiver acima de 0,7 é considerado um desempenho satisfatório.

Qualidade de estudos de testes diagnósticos/ rastreamento

Para avaliação da qualidade metodológica dos testes diagnósticos e de rastreamento, existem guias de orientação, como o QUADAS (*Quality Assessment of Studies of Diagnostic Accuracy included in Systematic Reviews*), instrumento de avaliação da qualidade metodológica de testes diagnósticos incluídos em uma revisão sistemática, atualmente atualizado para a versão QUADAS 2. Este guia consiste em um *checklist* com quatro domínios chaves para avaliação: seleção de paciente, teste índice, padrão de referência, fluxo e período de seguimento. Além do QUADAS, outros instrumentos com a mesma finalidade, o STARD (*Standards for Reporting Diagnostic accuracy*) e TRIPOD (*Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis*), disponibilizados no Equator Network (www.equator-network.org/reporting-guidelines), podem ser utilizados. Estes instrumentos oferecem modelos de fluxogramas que permitem esquematizar o fluxo da seleção dos pacientes, composição dos subgrupos e, além disso, incluir os resultados do teste índice. O CASP (*Critical Appraisal Skills Programme*) (casp-uk.net/casp-tools-checklists) também oferece um guia para avaliação de estudos de testes diagnósticos, além de outros *checklist* para estudos de revisão sistemática e estudos observacionais.

O instrumento STARD, criado em 2003 e atualizado em 2015, auxilia na padronização do registro dos testes diagnósticos, através dos seguintes componentes:

1. título;
2. resumo;
3. introdução;
4. metodologia, na qual, se avalia o delineamento do estudo, participantes, métodos do teste índice e do teste padrão;
5. resultados, apresentados no diagrama de fluxo, dados de base, características clínicas dos pacientes, severidade da doença, alternativas diagnósticas, intervalos de tempo, a tabela cruzada com os resultados em comparação com o teste padrão, precisão estimada e eventos adversos;
6. discussão, que inclui as limitações, generalização dos

resultados e interpretação e;

7. outras informações, como registros, protocolo e financiamento.

Adicionalmente, a ferramenta TRIPOD, criada em 2015, fornece recomendações para reportar modelos de testes para diagnóstico, compostas por seis componentes:

1. título e resumo;

2. introdução;

3. metodologia;

4. resultados, no qual são apresentados o diagrama de fluxo, especificações e desempenho do modelo proposto;

5. discussão, incluindo as limitações, generalização dos resultados e interpretação e;

6. outras informações como, registros, protocolo e financiamento.

Avaliando estudos de testes diagnósticos e rastreamento

Existem três perguntas que orientam a análise crítica de estudos de testes diagnósticos ou de rastreamento: o estudo é válido? O teste índice distingue com precisão as pessoas doentes e não doentes? Qual a aplicabilidade do teste índice?

Validade do estudo

Se for constatado que o estudo avaliado apresenta falhas em algum dos aspectos abaixo, é necessário considerar se tais falhas da pesquisa invalidam os resultados. Para avaliar se um estudo é válido, as seguintes perguntas devem ser realizadas:

- Houve uma comparação independente e em caráter cego com um padrão de referência (padrão-ouro)?

- O viés de investigação foi evitado? Ou seja, todos os que realizaram o novo teste diagnóstico também realizaram o padrão-ouro e vice-versa?

- O viés de expectativa foi evitado? Ou seja, avaliação foi submetida a mascaramento?

Os pacientes do estudo devem ser submetidos tanto

ao teste índice como ao teste de referência, a fim de que seja confirmado ou rejeitado os achados do teste índice. Isso significa que os participantes da pesquisa devem ser submetidos aos dois procedimentos diagnósticos, aquele que está sendo testado e o padrão-ouro, reconhecido como prova de que o paciente tem ou não uma determinada doença. Logo, o padrão de referência deve ser aplicado em todos os pacientes em estudo.

Além disso, os resultados de um procedimento não podem ser conhecidos pelo pesquisador que aplica ou interpreta o teste que está sendo comparado. Isto é, para que o viés de investigação seja evitado, os pesquisadores, que avaliam os resultados do teste índice, devem estar cegados para os resultados do padrão de referência. A precisão do teste pode ser superestimada, se realizado o teste índice primeiro em indivíduos sabidamente doentes e depois em indivíduos não doentes, em vez de proceder os testes índice e de referência no mesmo grupo de pessoas, sem saber se elas têm ou não a doença que está sendo investigada, especialmente se o teste de referência estiver aberto à interpretação subjetiva. Esses dois elementos são fundamentais para que sejam afastados os vieses na interpretação, conscientes ou inconscientes, dos pesquisadores.

Por fim, verifica-se se o padrão-ouro utilizado na pesquisa realmente merece este nome. Para muitos problemas, não há um teste diagnóstico padrão-ouro absoluto, que defina com certeza se os participantes estão doentes ou não. Assim, os pesquisadores podem necessitar desenvolver e justificar uma combinação de critérios, em relação aos quais o novo teste será avaliado. Um ponto específico a verificar é que o teste que está sendo validado, não esteja englobado também na definição do padrão-ouro.

As seguintes perguntas devem ser respondidas:

1. o teste índice foi avaliado em um espectro apropriado de pacientes (como aqueles que veríamos na prática)? O estudo deve incluir indivíduos de todo o espectro da doença, com sintomas de manifestações precoces, bem como sintomas mais graves e/ou pessoas com outras alterações que são comumente confundidas com a doença de interesse. Os

pacientes em estudo devem apresentar os achados clínicos, comumente presentes na doença em questão, bem como as características clínicas presentes em doenças que entram no diagnóstico diferencial. Estudos que aplicam testes diagnósticos, comparando casos clínicos atípicos muito específicos, ou utilizando pacientes assintomáticos, não são relevantes. Assim como, pesquisas que incluem apenas pessoas com sintomas comuns e assintomático, também, não são muito úteis. Portanto, a fim de evitar uma amostra tendenciosa, torna-se necessário estabelecer critérios específicos de elegibilidade, além de incluir participantes com diferentes espectros de idade, gênero, sintomas e/ou gravidade da doença.

É muito comum que estudos omitam o espectro de participantes. Definir tanto a variedade de participantes quanto o espectro da doença a ser incluído é essencial para que os valores das diferentes características do teste mereçam ser citados, ou seja, para que possam ser aplicáveis em outras situações, favorecendo a validade externa;

2. o padrão de referência foi aplicado independentemente do resultado do teste índice? Há situações em que, se o indivíduo tiver um resultado de teste índice negativo, os pesquisadores, às vezes, não realizarão o teste padrão para confirmar o resultado negativo, especialmente se o teste é invasivo ou arriscado, pois isso pode ser antiético. Nestes casos, os pesquisadores empregam um padrão de referência alternativo para provar que o indivíduo não tem a alteração de interesse, que é o acompanhamento de longo prazo para avaliar que não há efeitos adversos associados à alteração alvo presente sem qualquer tratamento. Isto porque é necessário confirmar a precisão do teste índice, ou seja, que o resultado seja negativo em alguém sem a doença;

3. o teste foi validado em um segundo grupo independente de pacientes? Quando um novo teste diagnóstico ou de rastreamento é avaliado, há o risco de que os resultados da avaliação inicial sejam causados por outros fatores, por exemplo, alguma característica específica dos grupos de

pacientes incluídos no estudo, como sintomas avançados da doença. O primeiro estudo com o teste diagnóstico serve para se fazer uma predição da sua acurácia. Entretanto, a confirmação dos níveis semelhantes de desempenho com a aplicação em uma segunda amostra, independente, de pacientes permite a estimativa real da acurácia do teste diagnóstico em estudo.

Assim, para provar que os resultados são confiáveis e replicáveis, o novo teste deve ser avaliado em um segundo grupo de pacientes independentes. Visto que se os resultados desse segundo grupo de participantes forem semelhantes aos resultados do primeiro grupo, poderemos ter certeza sobre a precisão do teste. Porém, se nenhum estudo de conjunto de testes tiver sido realizado, é necessário cuidado nas conclusões.

Precisão do teste índice

A precisão do teste índice refere-se a sua capacidade de distinguir as pessoas doentes das não doentes, para isto, o resultado do teste índice deve ser analisado. Para avaliar se um teste índice distingue com precisão doentes e não doentes, as seguintes perguntas devem ser consideradas:

1. como avaliamos os resultados do teste? Os resultados do teste índice são avaliados através da análise dos parâmetros fornecidos pela tabela de contingência 2×2 : sensibilidade, especificidade, VPP, VPN, RV+ e RV-. Não há valores absolutos para determinar a validade de um teste de rastreamento, pois o que conta como aceitável depende do problema que está sendo rastreado;
2. o teste mostrou ser reprodutível tanto intra- como interobservadores? Alguns fatores que podem contribuir para a baixa reprodutibilidade de um teste diagnóstico são a precisão técnica do equipamento e a variabilidade do observador. Frente ao mesmo resultado de um exame para interpretar, duas pessoas concordarão somente em uma proporção dos casos. Esta reprodutibilidade entre observadores, geralmente, é expressa pelo índice Kappa;

3. os intervalos de confiança para sensibilidade e especificidade do teste foram fornecidos? O intervalo de confiança expressa a possível variação de resultados dentro dos quais se situa o valor verdadeiro. Quanto maior o tamanho da amostra, menor o intervalo de confiança, de modo que é particularmente importante verificar os intervalos de confiança se o artigo que você estiver lendo descrever um estudo realizado em uma amostra relativamente pequena;

4. qual o tamanho da amostra do estudo? É importante o cálculo do tamanho amostral, pois todos os parâmetros aferidos do teste índice dependem da prevalência da doença. O cálculo do tamanho da amostra, segue os mesmos princípios estabelecidos em estudos de prevalência, sendo necessária: (1) estimativa da proporção esperada da positividade (doentes) na população; (2) amplitude do intervalo de confiança que se deseja e; (3) definição do intervalo de confiança, geralmente de 95%;

5. foi estabelecido um ponto de corte? Se o teste fornece resultados não dicotômicos (contínuos), ou seja, se o teste fornece um valor numérico e não um resultado qualitativo, deve-se determinar quais valores do resultado do teste serão considerados anormais. O problema de definir uma faixa de normalidade para uma variável contínua é determinar o limiar de transição. Simplesmente, não há um ponto de corte preciso entre o normal e o anormal, devido a zona de transição ou "zona cinzenta".

É possível usar os resultados de um estudo de validação para desenhar uma série completa de tabelas de contingência 2 X 2. Cada tabela usaria uma definição diferente de resultado anormal do teste índice (ponto de corte) para classificar os pacientes como normais ou não. A partir dessas tabelas, poderíamos gerar diferentes razões de verossimilhança. É por isso que a abordagem recomendada nestes casos é uma combinação de diferentes exames. Embora qualquer teste possua limites imprecisos entre normal e anormal, combiná-los pode afinar o foco diagnóstico.

Aplicabilidade do teste índice

Por fim, para estabelecer se o teste índice é aplicável na prática clínica, as seguintes questões devem ser consideradas:

1. o teste é potencialmente relevante para a prática? Para responder a esta pergunta, verifique se o método considerado é útil no manejo do paciente, se o custo é exequível, se o novo método realmente é vantajoso para substituir os testes existentes e se é capaz de identificar um transtorno tratável;

2. como aplicar o teste índice em um paciente? Tendo encontrado um estudo de teste diagnóstico ou de rastreamento válido e decidido que sua precisão é suficientemente alta para torná-lo uma ferramenta útil, alguns pontos precisam ser considerados ao aplicar o teste a um paciente: o teste está disponível, acessível e preciso em nosso cenário? Quão atual é o estudo analisado? As evidências mudaram desde a publicação do estudo? Os pacientes do estudo são semelhantes ao paciente em questão?

3. a probabilidade pós-teste afetará o manejo do paciente? Recomenda-se ponderar se os resultados do teste alteram a conduta e evitam testes adicionais, ou seja, se o método descarta a condição alvo. Ou então, avaliar se o teste faz um diagnóstico da condição de interesse, propiciando opção de tratamento adequado. Além disso, certifique-se de que os pacientes estão dispostos a se submeter ao método, e se os resultados do exame o ajudarão a alcançar seus objetivos.

Referências

Bonita, R.; Beaglehole, R.; Kjellström, T. Epidemiologia básica. 2. ed. São Paulo: Santos, 2010-2019. 213 p.

Bossuyt PM. Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem. 2003 Jan;49(1):1-6. doi: 10.1373/49.1.1. PMID: 12507953.

Bossuyt PM. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015 Oct 28;351:h5527. doi: 10.1136/bmj.h5527. PMID: 26511519; PMCID: PMC4623764.

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015 Jan 6;162(1):55-63.

Ferreira & Patino. Entendendo os testes diagnósticos: parte 2. *J Bras Pneumol*. 2017;43(6):408-408.

Ferreira & Patino. Entendendo os testes diagnósticos. Parte 1. *J Bras Pneumol*. 2017;43(5):330-330.

Ferreira & Patino. Entendendo os testes diagnósticos. Parte 3. *J Bras Pneumol*. 2018;44(1):4-4.

Greenhalgh, T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.

Whiting, P. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 3, 25 (2003).

Estudos observacionais: coorte, caso-controle e transversal

*Caroline Darski
Michele Strelow Moreira
Edison Capp
Otto Henrique Nienov*

Os estudos observacionais limitam-se à observação da realidade para geração de conhecimento, em que o pesquisador não tem controle sobre as variáveis de estudo, o que aumenta o risco de vieses. Os estudos observacionais podem ser descritivos e analíticos. Um estudo descritivo, como o relato de caso e a série de casos, limita-se a descrever a ocorrência de uma doença em uma população. Os estudos analíticos tentam quantificar a relação entre dois fatores, ou seja, o efeito de uma exposição sobre um desfecho e incluem os estudos do tipo coorte, caso-controle e transversal.

Estudos de coorte

O termo coorte tem origem no latim "cohors", que identificava um décimo das legiões de soldados romanos que marchavam juntos de um ponto geográfico a outro durante as campanhas bélicas do Império. Em epidemiologia, o paralelismo com as coortes romanas é porque acompanhamos grupos de indivíduos por um período determinado para avaliar desfechos clínicos.

O estudo de coorte, também chamado de estudo longitudinal, *follow-up*, de seguimento, sequencial ou de incidência, é um modelo de estudo do tipo primário, analítico e observacional, no qual os indivíduos são classificados (ou selecionados) segundo o estado de exposição, em exposto e não exposto, sendo seguidos para avaliar a incidência da doença em determinado período de tempo. O investigador limita-se a observar e analisar a relação existente entre a presença de fatores

de risco ou características e o desenvolvimento de doenças, em grupos da população. Sendo a unidade de observação o indivíduo, o acompanhamento permite detectar as mudanças que ocorreram em todos os participantes do estudo.

Modelo conceitual do estudo de coorte

O modelo conceitual de um estudo de coorte é relativamente simples. O estudo de coorte inicia com um grupo de pessoas livres de doença com alguma característica em comum (por exemplo, uma coorte de nascimento compartilha o mesmo ano ou período de nascimento, uma coorte de fumantes tem em comum o hábito de fumar, uma coorte de vegetarianos compartilha a mesma dieta, etc.), em que dois (ou mais) subgrupos de pessoas são selecionados com base nas diferenças em sua exposição a um agente específico (fator(es) de exposição), que divide o grupo em exposto (que possuem o fator de risco ou característica de interesse) e não exposto (que não possuem o fator de risco ou característica de interesse), e são acompanhados para observar quantos em cada grupo desenvolvem determinada doença, complicação ou outro desfecho sob investigação (incidência ou ocorrência).

O período de acompanhamento em estudos de coorte geralmente é medido em anos, e pode ser prospectivo (Figura 1) ou retrospectivo (Figura 2). Portanto, o parâmetro a ser estudado é a presença ou não da doença. Compara-se a proporção dos que ficaram doentes dentre os expostos, e a proporção dos que ficaram doentes entre os não expostos. As variáveis de interesse são especificadas e medidas e a coorte inteira acompanhada com o objetivo de ver se o surgimento de novos casos de doença (ou outro desfecho) difere entre os grupos, conforme a presença ou não de exposição. Em virtude de os dados coletados fazerem referência a diferentes pontos no tempo, os estudos de coorte são longitudinais.

Por exemplo, uma coorte de mulheres no climatério é acompanhada anualmente, por vários anos, para observar a incidência de infarto agudo do miocárdio (IAM) naquelas que receberam terapia de reposição hormonal (TRH) e nas que não receberam TRH (Figura 3).

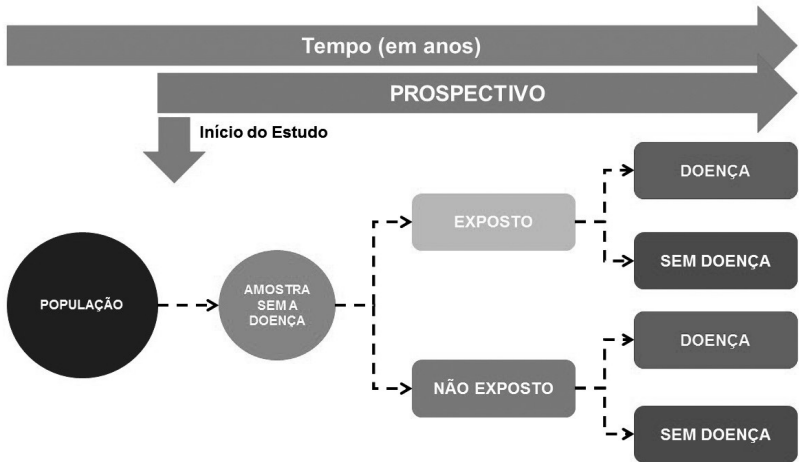


Figura 1. Modelo conceitual de um estudo de coorte prospectivo.

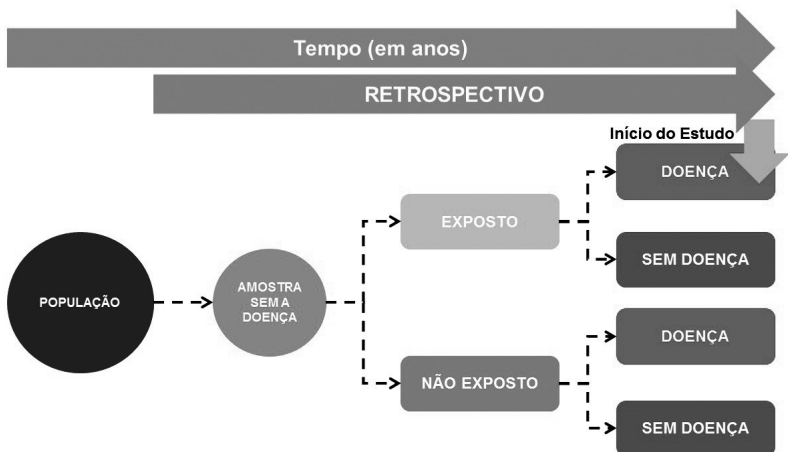


Figura 2. Modelo conceitual de um estudo de coorte retrospectivo.

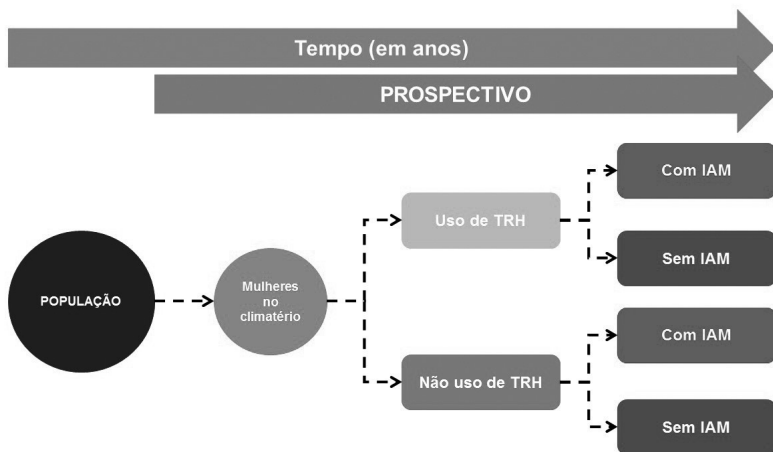


Figura 3. Exemplo de estudo de coorte prospectivo.

Fases do estudo de coorte

As fases principais de um estudo de coorte são as seguintes:

1. Identificação das pessoas saudáveis no início do estudo;
2. Montagem dos grupos de indivíduos expostos e não expostos;
3. Seguimento da coorte para avaliação da incidência da doença a ser estudada nos dois grupos e;
4. Comparação da incidência (risco) em cada coorte (exposto e não exposto).

Lembre-se que nos estudos de coorte, os pesquisadores observam o estado de exposição, que ocorre “naturalmente” (ou seja, o pesquisador não interfere ativamente). Em contraste, nos ensaios clínicos, os pesquisadores atribuem (experimentalmente) a exposição aos participantes do estudo.

Como fazer a seleção do grupo de comparação?

Os grupos ou subgrupos dos estudos de coorte devem ser tão semelhantes quanto possível em relação a outros fatores que podem estar relacionados à doença, exceto o fator de exposição

em investigação. Poucos estudos de coorte são bem-sucedidos na identificação de grupos ou subgrupos de sujeitos semelhantes em idade, proporção entre os sexos, nível socioeconômico, presença de doenças coexistentes, entre outros aspectos, com a única diferença sendo a sua exposição ao agente de interesse. Na prática, boa parte do “controle” nos estudos de coorte ocorre na fase de análise, em que é feito um ajuste estatístico complexo para diferenças basais em variáveis-chave.

Amostragem

Os critérios de elegibilidade para inclusão dos participantes na pesquisa devem ser estabelecidos no protocolo de pesquisa. Também deve-se determinar o tempo e a intensidade da exposição do grupo. É importante certificar-se de que, na seleção, o participante não apresente a doença de interesse (desfecho). A escolha do grupo de comparação (não expostos) deve obedecer aos mesmos critérios empregados para seleção do grupo de expostos.

Se os participantes são selecionados através de uma amostra da população, classificados de acordo com a presença ou não da exposição, então o estudo é de comparação interna. O uso de grupos de comparação interna aumenta a probabilidade dos participantes de ambas as coortes pertencerem a subgrupos similares da população e implica que os grupos serão submetidos aos mesmos procedimentos durante o tempo de seguimento, tendo a mesma chance de ter a doença detectada.

Quando houver comparação externa, o grupo não exposto é selecionado como uma amostra da população geral de onde o grupo exposto foi identificado e os dois grupos podem diferir em relação a outros fatores de risco. O maior problema metodológico associado à um estudo de coorte é o de manter o seguimento (aderência) da população selecionada.

O que acontece com alguém que tem a doença?

Os estudos de coorte geralmente se concentram na incidência de casos da doença que ocorrem durante o seguimento. Para atingir esse objetivo, os pesquisadores geralmente excluem indivíduos que têm a doença prevalente no início de um estudo

de coorte. Um tipo especial de estudo de coorte também pode ser usado para determinar o prognóstico de uma doença.

Desta forma, podemos ter um estudo prognóstico (Figura 4), que consiste no acompanhamento de um grupo de pacientes que tenham sido diagnosticados como apresentando um estágio inicial da doença ou um exame de rastreamento positivo por um período de tempo, acompanhado repetidamente para detectar a persistência da doença, o que possibilita o estudo evolutivo da enfermidade. Pode-se, então, dispor de vários dados: aparecimento de novos sintomas e sinais no decorrer do tempo, com sua frequência; dados relacionados ao prognóstico; etc.

Neste caso, ao invés da incidência, estaremos avaliando a persistência da doença. Os pacientes devem ter diagnóstico recente, serem seguidos por tempo suficiente e submetidos apenas ao tratamento padrão ou, quando possível, a nenhum tratamento. No entanto, devido ao custo elevado e ao longo tempo necessário de observação, na maioria das vezes, opta-se por estudos retrospectivos baseados em dados armazenados em prontuário, que tem a vantagem de serem baratos e possuírem tempo longo de observação. No entanto, a principal desvantagem é que as informações clínicas podem ser incompletas e o estudo acaba dependendo da qualidade do registro das informações.

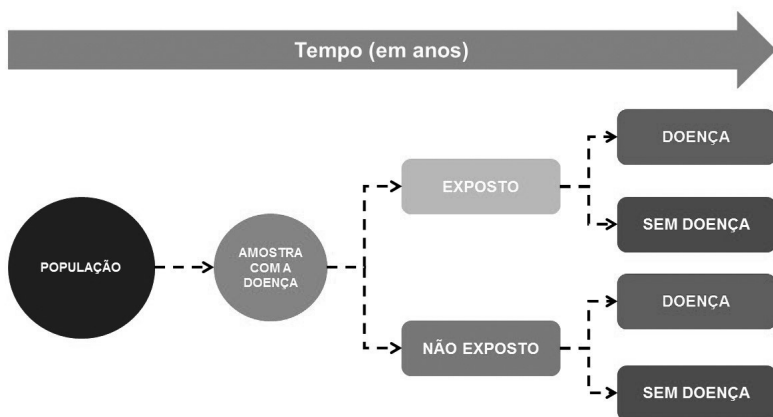


Figura 4. Modelo conceitual de um estudo de prognóstico.

Exposição, desfecho e fatores de confusão

Frequentemente, os fatores de exposição estudados são de risco para a doença (por exemplo, tabagismo, hipertensão) e os desfechos são as doenças desenvolvidas (por exemplo, câncer de pulmão, IAM). Como resultado, indivíduos expostos e não expostos podem variar em função de outras características. Os fatores de confusão correspondem a outros fatores, que não a exposição, que podem influenciar os resultados do estudo (por exemplo, cafeína).

Um dos aspectos importantes nos estudos de coorte é a classificação dos indivíduos nos grupos expostos e não expostos. As técnicas usadas para medir a exposição variam: entrevistas, questionários, exames laboratoriais, avaliações clínicas, prontuários médicos, etc. Os participantes podem ser classificados em várias categorias da exposição, incluindo a ausência à exposição. Se os participantes não forem corretamente classificados, o estudo fica seriamente comprometido.

Durante o desenvolvimento do estudo podem ocorrer mudanças na exposição aos fatores de risco, como mulheres modificarem os métodos contraceptivos utilizados. Também a própria participação no estudo pode levar os participantes a alterarem sua exposição em consequência das informações recebidas, como por exemplo, os participantes de um estudo sobre nutrição, questionados sobre hábitos alimentares, podem alterar sua alimentação. Quando isto ocorre, o investigador não tem controle em relação à estas mudanças; os participantes são livres para escolher e em muitos casos as modificações são desejáveis. Entretanto, é importante registrar durante o estudo todas as mudanças que ocorrerem, suas razões e os diferentes períodos de exposição, para consideração na análise de dados.

A avaliação do evento de interesse, da mesma forma, pode ser coletada em exames clínicos, laboratoriais, entrevistas, questionários, revisão de prontuários médicos, revisão da declaração de óbito, realizadas periodicamente durante o desenvolvimento do estudo. A periodicidade da avaliação do desfecho depende primariamente da condição que está sendo investigada. É importante obter informações de outras características dos grupos de estudo, tais como idade, sexo e

outros fatores de interesse, com a finalidade de identificar outras variáveis que possam estar relacionados à doença ou à exposição. O estudo pode ser limitado à qualidade das informações.

Os procedimentos para identificação de doenças deverão ser os mesmos para os expostos e os não expostos e, sempre que possível, para evitar vícios, o diagnóstico clínico ou laboratorial da doença deve ser realizado por profissionais que desconheçam a exposição do participante. Quando são utilizados dados de prontuários para determinar a ocorrência de doenças, é necessária atenção especial, pois as informações coletadas de diferentes hospitais podem não ser padronizadas, os critérios diagnósticos utilizados podem variar e alguns prontuários podem ser mais completos do que outros. Os expostos podem buscar mais os serviços de saúde comparados aos não expostos. Neste caso, a ocorrência da doença pode ser superestimada. A definição de caso deve basear-se em testes sensíveis e específicos, principalmente para aquelas doenças subclínicas ou assintomáticas.

Os fatores de confusão são caracterizados como uma terceira variável que:

1. está causalmente associada ao evento;
2. está associada com a exposição;
3. está diferencialmente distribuída entre os grupos (expostos e não expostos) e;
4. não é parte do mecanismo causal que liga a exposição e o evento.

A confusão pode resultar em uma superestimação ou subestimação da verdadeira associação.

Medição dos dados

A partir do momento em que os pesquisadores decidem sobre as variáveis e os desfechos específicos para o estudo, estes devem usar os métodos mais acurados possíveis para medir essas características, disponibilizando-os para todos os participantes da pesquisa. Considerações importantes na medição dos dados do

estudo são a validade das medições, o tempo entre as medições e a disponibilidade de medidas uniformes para a população em estudo.

A validade de uma medição refere-se a quão perto os dados medidos representam os dados verdadeiros, devendo ser a mais próxima da realidade possível. O tempo entre as medições da exposição e do desfecho dependerá do conhecimento prévio sobre o desfecho, ou seja, da relação entre a exposição e a doença. Deve-se considerar se a exposição pode mudar durante o estudo, como por exemplo, o uso de algum medicamento. Por isso é importante o acompanhamento dos participantes. Os estudos de coorte não são ideais para o estudo de exposições muito recentes ou muito antigas. Exposições antigas podem mudar ao longo do tempo, obscurecendo a sua relação com o resultado de interesse. Exposições muito recentes podem confundir a relação presumida temporal entre a exposição e o desfecho.

Temporalidade nos estudos de coorte

Os estudos de coorte podem ser classificados como concorrentes (prospectivos ou clássicos) ou não concorrentes (retrospectivos). Os termos prospectivo e retrospectivo geralmente se referem a quando os dados do estudo foram coletados em relação ao pesquisador. Um estudo prospectivo envolve a coleta de novos dados, muitas vezes, com o propósito de abordar uma questão específica. A distinção entre os estudos retrospectivos e prospectivos é essencialmente descritiva, e não traz geralmente impacto no plano de análise ou nos resultados do estudo.

Nas coortes prospectivas (Figura 1), o investigador define, claramente, um critério para considerar se o sujeito possui ou não o desfecho de interesse e, então, inicia-se o estudo com um grupo sem a doença. Nos estudos concorrentes, a exposição pode (ou não) já ter ocorrido, sendo que a duração e intensidade da exposição podem influenciar o evento de interesse, mas o desfecho ainda não ocorreu. Quando prospectivo, geralmente, é caro e de longa duração. Como vantagens, o pesquisador tem a possibilidade de avaliar diretamente a exposição, minimizando vieses, e pode obter dados sobre fatores de confusão.

Nos estudos não concorrentes (Figura 2), todas as informações sobre a exposição e o desfecho já ocorreram antes do início do estudo. O estudo de coorte retrospectivo (coorte histórica) é vantajoso em relação ao custo (mais barato e rápido), porém está mais sujeito a vieses, principalmente de seleção, pois todos os dados sobre exposição e efeito (doença) foram coletados antes do início do estudo, viés de informação (depende de disponibilidade de registros) e a incapacidade para controlar variáveis de confusão (falta de informação). Desta forma, o desenho da coorte, as medições basais, o acompanhamento e os desfechos ocorreram todos no passado. Esse tipo de estudo só é possível se houver dados adequados disponíveis sobre os fatores e os desfechos. Os estudos não concorrentes raramente podem ser conduzidos em uma amostra geral da população. Para que isto ocorra, é necessário que o investigador tenha acesso às informações sobre a exposição ao fator de risco no passado, ou seja, no início da investigação. Somente se estas informações estiverem disponíveis será possível identificar amostras de participantes expostos e não expostos e acompanhar estes indivíduos selecionados.

Duração de um estudo de coorte

Dependendo do desfecho e da exposição do estudo, o tempo de seguimento dos grupos expostos e não expostos pode variar de horas a décadas. Estima-se a duração e a intensidade desta exposição através da utilização de instrumentos como questionários ou entrevistas, prontuários médicos ou testes diagnósticos.

Quando os estudos de coorte são indicados?

O estudo de coorte, pelas suas características, dá informações valiosas sobre causas das doenças, sobre sua incidência, evolução do quadro clínico (por exemplo, evolução de portadores do HIV) e prognóstico (por exemplo, marcadores tumorais e evolução de câncer). Os estudos de coorte fornecem a melhor informação sobre a etiologia das doenças (por exemplo, associação entre tabagismo e câncer de pulmão) e a medida mais direta do risco de desenvolvê-la.



Figura 5. Aplicação dos estudos de coorte.

Incidência, risco relativo e risco atribuível

A incidência é medida pela proporção de um grupo inicialmente livre de uma condição clínica, e que a desenvolve depois de um período determinado de tempo. Ou seja, é o estudo de casos novos ou de desfechos novos dos casos existentes de uma doença específica, que ocorrem numa população que não os apresentava. Detectam-se os novos casos ou desfechos, ao longo do tempo, mediante a realização de exames periódicos. Na prática, utiliza-se a medida de incidência como uma estimativa de risco. Independentemente do número de coortes, deve-se primeiro calcular a incidência da doença em cada grupo, ou como proporção de incidência ou como taxa de incidência. Isso deve ser feito levando-se em conta o tempo em que cada participante foi seguido no estudo, dando como resultado para cada indivíduo um valor de pessoa-tempo. O tempo de seguimento no denominador da taxa de incidência padroniza as taxas da doença por tempo.

O risco relativo (RR) é definido como a incidência (proporção ou taxa) em uma coorte dividida pela incidência na coorte de referência, ou seja, é a razão entre a incidência nos expostos (probabilidade do indivíduo ficar doente dado que foi exposto ao fator de interesse) pela incidência nos não expostos (probabilidade de ficar doente dado que não foi exposto ao fator

de interesse). Em outras palavras, é o número de eventos no grupo exposto dividido pelo número total de expostos (incidência nos expostos) e este dividido pelo número de eventos no grupo não exposto dividido pelo número total de não expostos (incidência nos não expostos) (Figura 6).

$$RR = \frac{P_{\text{exposto}}}{P_{\text{controle}}} = \frac{\left[\frac{a}{(a+b)} \right]}{\left[\frac{c}{(c+d)} \right]}$$

Figura 6. Cálculo do risco relativo.

Os dados de indivíduos expostos e não expostos, doentes e não doentes, para calcular a incidência, podem ser reunidos em uma tabela de contingência 2 x 2 (Figura 7).

		DESFECHO (doença)	
		DOENTE (presente)	NÃO DOENTE (ausente)
FATOR	EXPOSTO (presente)	a	b
	NÃO EXPOSTO (ausente)	c	d

Figura 7. Tabela de contingência 2 x 2 do estudo de coorte.

O RR representa quantas vezes mais o indivíduo exposto está sujeito a desenvolver a doença em relação ao não exposto. Ou seja, a magnitude do RR reflete a força da associação fator de risco-doença. Assim temos:

- RR > 1,0: maior será a convicção de que o fator em estudo está associado ao desfecho (doença) de interesse. Indica associação positiva ou risco aumentado entre os expostos ao fator estudado em relação aos não expostos;
- RR < 1,0: maior será a convicção de que o fator em estudo

é protetor ao desfecho (doença) de interesse. Indica que há uma associação inversa ou risco diminuído entre os expostos ao fator estudado, também chamado de fator de proteção;

- RR = 1,0: a incidência entre os expostos e não expostos é igual, ou seja, o fator de exposição em estudo não tem relação com a doença. Indica que a taxa de incidência da doença nos grupos exposto e não exposto são idênticas, indicando que não há associação observada entre exposição e desfecho.

Lembre-se também de verificar o intervalo de confiança de 95% (IC95%) do RR calculado. O IC95% corresponde a faixa de valores entre os quais podemos ter 95% de confiança de que o resultado estará incluído. A significância do RR é calculada pelo IC:

- se o IC contiver o valor 1 (um inteiro), significa que o fator de exposição parece não acrescentar risco nem proteção para o desfecho (doença);

- se o IC não contiver o valor 1 (um inteiro), o fator de exposição acrescenta risco ou é um fator de proteção para o desfecho (doença).

O risco atribuível na população (RAP), também chamado de “diferença de risco” ou “excesso de risco”, é definido como a incidência (proporção ou taxa) em uma coorte diminuída a incidência do outro grupo. Se a exposição de interesse for a causadora da doença, o risco atribuível descreve a quantidade de risco adicional ou extra que se deve à exposição.

Vantagens e desvantagens dos estudos de coorte

Dentre as vantagens dos estudos de coorte, pode-se destacar a capacidade de discernir as relações temporais entre a exposição e o desfecho pelo fato de a exposição preceder o desfecho. Ao contrário dos estudos transversais, estudos de coorte podem revelar relações temporais entre a exposição e o desfecho desde que uma quantidade razoável de tempo tenha decorrido

entre a medição da exposição e a ocorrência do resultado. A existência de uma relação temporal reforça a evidência para a exposição ser uma causa possível da doença.

Os estudos de coorte podem ser utilizados para avaliar desfechos múltiplos. Os pesquisadores podem estudar mais de um desfecho desde que os sujeitos em estudo estejam livres de cada desfecho de interesse quando o estudo iniciou, assim como permitem a avaliação de várias exposições (coortes de população geral ou de grupos populacionais restritos). Também permitem o cálculo direto das medidas de incidência nas coortes de expostos e não expostos e o conhecimento da história natural da doença. Além disso, o status do desfecho não influencia a medida do status de exposição ou seleção de indivíduos na coorte prospectiva e não apresentam problemas éticos quanto à decisão em expor os participantes a fatores de risco ou tratamento como ocorre nos ensaios clínicos.

Embora conceitualmente simples, os estudos de coorte são bastante caros porque podem requerer longos períodos de acompanhamento, visto que a doença pode ocorrer após uma exposição prolongada. Os estudos de coorte limitam-se por serem estudos observacionais e, portanto, estão sujeitos a fatores de confusão. Ou seja, outros fatores que estão relacionados com a exposição de interesse podem explicar algumas ou todas as associações que são observadas. Os sujeitos expostos podem ter acompanhamento diferenciado dos não-expostos. O estudo de coorte pode ser ineficiente para o estudo de doenças raras (exige grande amostra) ou aquelas com longos períodos de latência entre a exposição e o desfecho. Geralmente são caros e difíceis de operacionalizar (em estudos etiológicos) e a perda de participantes ao longo do seguimento pode comprometer a validade dos resultados.

Vieses em estudos de coorte

Nos estudos de coorte, dificilmente ocorre viés de seleção, pois os grupos são determinados de acordo com o fator de exposição e não há desfecho algum no momento da seleção. No entanto, pode ocorrer viés do voluntário (efeito do trabalhador saudável).

É comum ocorrer viés de perda de seguimento, pois alguns indivíduos expostos a determinado fator de risco podem evoluir para óbito, não comparecendo ao seguimento. Essa perda de acompanhamento pode estar relacionada tanto com a exposição como com o desfecho. Mesmo com perdas menores, representa um problema e é preciso verificar se ocorre perda diferencial entre os grupos exposto e não exposto. A perda de seguimento também pode ocorrer por migração (mudança de endereço e/ou contato telefônico), falta de aderência ou desistência (retirada do consentimento). Nos estudos de coorte ocorre o efeito de não participantes, em que os não participantes podem diferir dos participantes em níveis de motivação e atitudes em relação à saúde, assim como a fatores de risco.

O viés de informação em estudos de coorte são mais frequentes no momento de aferir o desfecho. Se os investigadores conhecem a exposição, podem ficar suggestionados ao aferir o desfecho. Refere-se a erros introduzidos durante a medição da exposição, dos eventos ou de outras variáveis, que influenciam a qualidade das medições nos grupos expostos e não expostos. Podem ser fontes de erro na aferição o observador (padronização), os participantes (viés de memória) e o instrumento (calibragem).

Redação de estudos de coorte

Para contornar alguns problemas, um grupo de pesquisadores das áreas de epidemiologia, metodologia científica e estatística desenvolveram um guia conhecido como STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*) (www.equator-network.org/reporting-guidelines/strobe).

A ferramenta STROBE foi desenvolvida para avaliar a qualidade de publicação de qualquer estudo epidemiológico do tipo coorte, caso-controle e transversal. Ela é composta por um *checklist* com 22 itens, desenvolvidos para recomendar tópicos essenciais em cada seção de um estudo com os desenhos abrangidos pela ferramenta.

Para estudos prognósticos em câncer, por exemplo, existe o REMARK (*REporting recommedations for tumor MARKer prognostic studies*) (www.ncbi.nlm.nih.gov/pmc/articles/PMC2361579/).

Lembre-se que estes *checklists*, além de serem utilizados como guias para redação dos manuscritos, também são excelentes para avaliar criticamente estudos observacionais.

Análise crítica de estudos de coorte

Algumas questões podem auxiliar na leitura crítica de estudos de coorte:

1. Existe viés de seleção no estudo? Foram incluídos apenas indivíduos sob risco para o desfecho? O fator de exposição foi específico, mensurável e definido por critérios objetivos? Os grupos expostos e não expostos eram semelhantes, exceto pelo fator de exposição? Todos os participantes do estudo (expostos e não expostos) devem estar sob risco de desenvolver o desfecho ou doença de interesse. Por exemplo, pacientes hysterectomizadas têm risco nulo de desenvolver câncer de útero, portanto, não devem ser incluídas em um estudo de coorte cujo desfecho seja câncer de colo do útero. Uma seleção inadequada de participantes contribui para redução na acurácia, ou seja, na capacidade de medir o que o estudo se propõe a aferir. Além disso, sujeitos sob suspeita de já serem portadores do desfecho em estudo não devem ser incluídos nos estudos de coorte, exceto nos de prognóstico.

De antemão, é necessário definir a condição de exposto e de não exposto, bem como o grau de exposição. O fator de exposição deve ser específico e mensurável. Por exemplo, em relação ao uso de fármacos vasoativos (FVAs), as coortes seriam 'uso de FVAs' (exposto) e 'não uso de FVAs' (não exposto). Em certos estudos, pode-se até mensurar o grau de exposição ao fator, dependendo do objetivo do estudo, como a dosagem de vasopressor infundido, nível de pressão positiva em ventilação mecânica não-invasiva, nível de pressão venosa central, contagem de leucócitos, tempo de protrombina, por exemplo.

Sob condições ideais, o grupo não exposto deve ser semelhante ao grupo exposto em todos os aspectos importantes, exceto pela ausência da exposição ao fator

de interesse. Os sujeitos de pesquisa incluídos no estudo devem ter características semelhantes à população-alvo, para justificar a generalização externa. Dessa maneira, o grupo não exposto irá revelar a taxa de eventos basal na população-alvo.

2. Existe viés de informação ou aferição? As informações sobre o desfecho foram obtidas da mesma maneira em ambos os grupos? A avaliação do desfecho foi realizada por um investigador cegado em relação ao fator de exposição? Depois de determinado tempo de seguimento será verificada a incidência do desfecho nos dois grupos (exposto e não exposto). É necessário que os desfechos sejam bem estabelecidos no planejamento do estudo. As informações sobre os desfechos podem ser oriundas de diversas fontes, como por exemplo para investigar taxa de mortalidade, geralmente é utilizado o atestado de óbito. Porém, para verificar desfechos não-fatais, como por exemplo o desenvolvimento de uma enfermidade, podem ser utilizadas diversas fontes como prontuário médico, exames laboratoriais, avaliação clínica, registro de alta hospitalar, etc. Se o desfecho de interesse for avaliado por equipamentos, esses devem ter as mesmas especificações e calibragem. O importante é que os desfechos sejam mensuráveis, comparáveis entre os grupos (exposto e não exposto) e definidos pelos mesmos critérios diagnósticos.

Idealmente, o avaliador dos desfechos não deve ter conhecimento (cegamento) em relação à exposição. A conhecimento da condição de exposição dos participantes pode influenciar o julgamento dos desfechos e gerar conclusões errôneas e enviesadas.

3. O seguimento dos sujeitos de pesquisa foi completo? Quais os esforços realizados para minimizar a perda de seguimento? A perda de seguimento foi semelhante nos dois grupos? A periodicidade do seguimento dos dois grupos de participantes deve ser previamente definida. Deve ser estabelecida a frequência dos exames

para diagnóstico do desfecho em estudo e da coleta de informações pertinentes para se evitar ao máximo as perdas de seguimento. As perdas de seguimento ou acompanhamento reduzem o poder estatístico dos testes e a precisão das estimativas. A descontinuidade do seguimento é a principal causa de perda de qualidade dos resultados nos estudos de coorte.

Mas, como evitar as perdas de seguimento? A melhor maneira de evitar perdas de seguimento é a seleção adequada da população de estudo, que deve ser avaliada quanto à possibilidade de continuidade. Além disso, existem outras maneiras de se assegurar o seguimento, como o levantamento de dados relacionados à identificação e também devem ser obtidos nomes e endereços de parentes ou de amigos, de maneira que o contato possa ser estabelecido mesmo com aqueles que se mudarem ou alterarem o contato telefônico. Contatos periódicos com a amostra de participantes, como o envio de lembretes, também evitam perdas de seguimento. No início do estudo, deve ser programada e informada ao participante a frequência das visitas, dos contatos telefônicos, das cartas ou questionários via e-mail.

Embora a perda de seguimento dos sujeitos de pesquisa diminua o poder e a precisão do estudo, a perda desbalanceada do ponto de vista quantitativo ou qualitativo entre os grupos (exposto e não exposto) é ainda mais prejudicial, pois evidencia um evento não aleatório. Se a probabilidade de perda está relacionada à exposição ou ao desfecho, pode haver viés de seguimento. Por exemplo, alguns sujeitos expostos a determinado fator de risco em estudo, que seriam acompanhados após a alta hospitalar, podem evoluir para óbito não comparecendo ao seguimento. A ausência desses sujeitos pode sugerir falsamente que o fator em estudo tenha menor relação com o desfecho.

4. Os potenciais fatores de confusão foram identificados e controlados na análise dos dados? Os investigadores

anteciparam e colheram informações sobre os potenciais fatores de confusão? Quais métodos foram utilizados para avaliar e controlar os fatores de confusão? Quando uma variável modifica a relação entre o fator de exposição e o desfecho é chamada de fator de confusão. Por exemplo, em um estudo de coorte sobre o efeito do uso de contraceptivo oral no desenvolvimento de IAM, pode-se encontrar o fator de confusão tabagismo. Usuárias de contraceptivo oral também podem ser tabagistas e, neste caso, o cigarro aumenta o risco para IAM. Embora os investigadores esperem avaliar o efeito do contraceptivo oral, estão de fato, avaliando o efeito de confusão do cigarro entre mulheres usuárias desse tipo de contracepção.

Por isso, é importante a identificação prévia dos prováveis fatores de confusão do estudo, definindo-os e incluindo-os nos critérios de elegibilidade. O viés de confusão, diferentemente dos vieses de seleção e informação, que são irreparáveis, são controláveis desde que informações pertinentes sejam coletadas, visando à homogeneidade entre os grupos do estudo. Pode-se controlar o viés de confusão por meio de:

- restrição na inclusão de pacientes que apresentem potenciais fatores de confusão;
- pareamento dos sujeitos incluídos nos grupos expostos e não expostos (ambos os grupos contêm pacientes com fatores de confusão na mesma proporção);
- análise em subgrupos, que pode ser considerada uma forma post hoc de restrição, feita durante a análise dos resultados por intermédio de métodos estatísticos;
- técnicas de análises múltiplas, onde modelos estatísticos examinam o potencial efeito de uma variável enquanto simultaneamente controlam o efeito de outros fatores.

A realização de análises em subgrupos (estratificadas) e análises múltiplas exige cautela, pois, a amostra inicialmente calculada pode não suportar tantas comparações. Aconselha-se manter no mínimo dez casos para cada fator acrescentado numa análise múltipla.

5. Os resultados obtidos pelo estudo são relevantes? A coorte teve amostra adequada? A análise estatística realizada foi adequada? A amostra normalmente é planejada para responder a uma questão de pesquisa: identificar associação entre o desfecho e o fator de interesse. Assim, os participantes serão classificados como desfecho positivo (doentes) ou desfecho negativo (não doentes) na presença do fator de exposição. O ideal é que todos os participantes sejam classificados corretamente.

Entretanto, às vezes erros são cometidos, ou seja, um sujeito não doente é dito como doente (erro tipo I) e um sujeito doente é avaliado como não doente (erro tipo II) (Figura 8). Assim, um erro tipo I é controlado no delineamento do estudo pelo nível de significância adotado (α), sendo sempre menor que o erro tipo II (β). O erro tipo II é controlado indiretamente pelo poder do teste, uma vez que o poder é representado por $1-\beta$. O poder do teste representa a capacidade do teste estatístico evidenciar a questão de pesquisa.

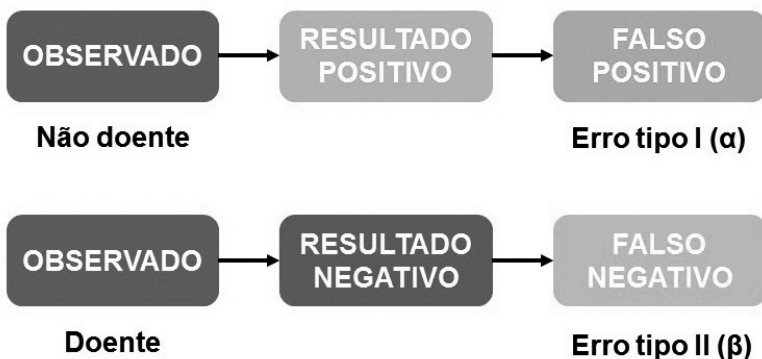


Figura 8. Tipos de erros em estudos de coorte.

Dessa forma, um estudo deve ter tamanho suficiente para evitar os dois tipos de erros: tipo I (afirmar que o fator de exposição ou de risco está associado com o desfecho e de fato não está) e tipo II (afirmar que o fator de exposição ou de risco não está associado com o desfecho e de fato está).

O número de sujeitos incluídos em um estudo de coorte depende da(o):

- incidência da doença nos grupos (exposto e não exposto);
- estimativa de RR associado à exposição;
- nível de significância desejado, normalmente fixado em 5% (que significa uma chance em 20 de se cometer o erro tipo I, isto é, de concluir que a exposição é relacionada ao evento de interesse quando na realidade isto não ocorre);
- poder do estudo, normalmente ficando em 80-90%. Quando especificado em 80%, teríamos 20% de probabilidade de se cometer o erro tipo II, ou seja, de concluir que a exposição não é relacionada ao evento de interesse, quando isto ocorre na realidade.
- e, ao número mínimo de pacientes calculado em cada grupo, acrescentam-se 20% a 30% para suprir eventuais perdas.

Dois aspectos importantes para o cálculo do tamanho de amostra são:

1. quanto maior o valor do RR, menor será o tamanho da amostra a ser estudada e, quanto menor o RR, maior tamanho de amostra;
2. os valores especificados para os parâmetros, quando modificados, alteram o tamanho da amostra.

O estudo de coorte pode ser inviabilizado pelo número de participantes necessário à sua realização. Cabe ao investigador decidir a viabilidade de sua execução, considerando os recursos financeiros, o tempo disponível e os riscos mínimos a serem detectados, que realmente tenham significado clínico ou biológico.

A maioria dos estudos de coorte tem como objetivo primário avaliar associação entre a incidência do desfecho em pacientes expostos e não expostos, sendo normalmente utilizado o teste Qui-quadrado quando não há fator de confusão. Também são calculados o RR, IC95% e RAP.

E, como considerar o tempo nos estudos de coorte? Pode-se realizar:

- análise pessoas-tempo: determinar as taxas em estudos que apresentam várias características (idade, sexo, grupo étnico, etc.) e diferentes tempos de seguimento que dificultam, ou mesmo impossibilitam, os cálculos necessários para a elaboração de uma tábua de vida. Considera-se tanto o número de pessoas que estão sendo observadas, quanto a duração dos períodos de observação, permitindo expressar em eventos por pessoa-tempo de observação, o período de tempo em que um número variável de pessoas é exposto ao risco deste evento. No entanto, limita-se pela premissa de que o risco da ocorrência de um evento por unidade de tempo é constante durante o período de observação;
- análise de sobrevida (tábuas de vida): cálculo da probabilidade de um evento por período de tempo especificado. Este método deve ser utilizado quando as premissas para o uso de pessoas-tempo não puderem ser satisfeitas.

E, quando realizar a estratificação? Quando não for indicado ajustar o efeito da variável de confusão no planejamento do estudo, utilizam-se métodos na análise de dados para neutralizar tal efeito. Os ajustes podem ser realizados através da estratificação dos dados pelas possíveis variáveis de confusão em múltiplas tabelas 2 x 2, pela técnica de Mantel-Haenszel, em que é feito o cálculo do RR nos estratos e do RR ajustado comparado ao RR não ajustado, e pelo modelo múltiplo logístico de Cox, que ajusta simultaneamente para o efeito de vários fatores de confusão.

6. os dados podem ser generalizados? Se os resultados forem relevantes e todas as possíveis causas de vieses forem descartadas, então, os dados gerados pela pesquisa devem ser avaliados em relação à validade externa, ou seja, a aplicabilidade clínica. A generalização dos resultados depende da similaridade das características basais dos pacientes incluídos no estudo e aqueles encontrados na prática. Se os pacientes preencherem os critérios de inclusão e exclusão estabelecidos pelo estudo de coorte, não há muita discussão quanto à aplicabilidade, ou seja, pacientes expostos àqueles fatores serão mais suscetíveis ao desfecho, necessitando de maior atenção.

Análise crítica de estudos de prognóstico

Algumas questões podem auxiliar na leitura crítica de estudos de prognóstico:

1. foi reunido um grupo representativo de pacientes no mesmo estágio da doença, preferencialmente inicial? A alocação ideal é aquela em que se inclui toda a população de pacientes com uma determinada doença, estudada a partir do seu início. Portanto, em estudos de prognóstico, deve-se observar como a doença foi definida e como os pacientes foram alocados.

A amostra deve representar as características da doença, da maneira que a encontramos na prática diária, mas principalmente, deve estar em seu curso inicial e no mesmo ponto em todos os pacientes alocados, o que é definido como uma coorte a partir da fase inicial da doença ou *inception cohort*. Se o objetivo for avaliar o prognóstico tardio de uma doença, e só nessa situação, que se deve alocar pacientes no mesmo ponto avançado de doença.

2. os pacientes foram acompanhados durante um tempo suficientemente longo e completo? Desfechos prognósticos devem ser avaliados de maneira que todos os pacientes tenham seguimento suficiente e completo para que possam manifestar o desfecho. O bom prognóstico

é constatado quando, após um tempo suficientemente longo, o evento adverso não está presente nos pacientes alocados. O estudo pode ser invalidado quando o seguimento dos doentes é interrompido antes que o evento adverso possa se manifestar. Não podemos estabelecer conclusões prognósticas baseadas em estudos cujo desfecho é analisado como um marcador intermediário, e nessa situação devemos buscar outra evidência. Seguir os pacientes ao longo do tempo tem como problema central o número de perda de seguimento dos pacientes que iniciaram o estudo. Lembre-se de avaliar o motivo da perda.

3. o critério de aferição do desfecho é objetivo e foi aplicado de maneira cegada? Os critérios de aferição do desfecho devem ser objetivos e claramente definidos no início do estudo. É preferível que o investigador que afere o desfecho seja cegado, ou seja, não conheça previamente as características clínicas e os fatores prognósticos dos pacientes, uma vez que poderá interferir na observação do desfecho, por mais que os critérios de aferição sejam objetivos.

4. se diferentes subgrupos prognósticos foram identificados: os mesmos foram ajustados com relação aos fatores prognósticos principais? Houve teste piloto em grupo independente de pacientes para validar os fatores prognósticos? Sabe-se que em grupo de pacientes estudados pode haver subgrupos com diferentes prognósticos. Muitas vezes esses subgrupos são claramente definidos, mas em outras situações não. Por exemplo, o risco de acidente vascular cerebral em pacientes com fibrilação atrial pode variar em função do tamanho do átrio esquerdo.

É necessário levar em consideração uma possível interação com outros fatores prognósticos relevantes, como sexo, idade, extensão da doença, comorbidades, etc. No exemplo citado, poderiam ser a hipertensão arterial e a função ventricular. A ocorrência de distribuição heterogênea desses fatores nos subgrupos pode provocar interação e influências mútuas que levem a um viés. Deve-

se avaliar a sessão do método e resultados em busca dessas características antes de se aceitar um prognóstico diferente para o subgrupo de interesse.

A análise estatística de subgrupos deve ser preditora e não explanatória. Não deve levar em consideração se o fator prognóstico é plausível no ponto de vista fisiológico, ou biologicamente sem sentido, ou ainda, quando é parte dos dados do paciente, mas semnexo causal com o desfecho.

Estudo de caso-controle

O principal objetivo dos estudos de caso-controle é verificar se há associação causal entre a exposição ao(s) fator(es) de risco e a presença ou ausência da doença em estudo, ou seja, do desfecho de interesse. Sendo assim, se o fator em estudo estiver associado com a doença, a proporção deste fator entre os casos será maior do que a proporção entre os controles.

O estudo de caso-controle, também chamado de caso-referência, é um estudo primário, analítico, observacional e longitudinal e, na maioria das vezes, retrospectivos, uma vez que o investigador busca, no passado, uma determinada causa (exposição) para a doença ocorrida. Entretanto, isso pode causar confusão porque os termos retrospectivos e prospectivos também são utilizados para descrever o tempo da coleta dos dados em relação ao momento atual. Neste sentido, um estudo de caso-controle pode ser tanto retrospectivo, quando os dados fazem referência ao passado, quanto prospectivo, quando os dados são continuamente coletados no decorrer do tempo.

Modelo conceitual do estudo de caso-controle

No estudo de caso-controle, os participantes portadores de uma determinada doença ou condição de saúde que o estudo objetiva avaliar (casos) em uma população (ou amostra representativa) são identificados e comparados com aqueles indivíduos que não são portadores da doença ou condição de saúde (controles). Ou seja, os grupos de estudo são definidos pelo desfecho, diferente do estudo de coorte cujos grupos são definidos pela exposição (Figura 9).

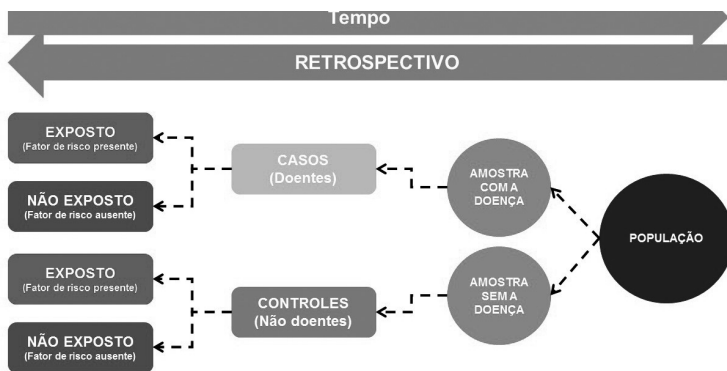


Figura 9. Modelo conceitual do estudo de caso-controle.

Este estudo, por ser controlado, permite a proposição de hipóteses, que podem ser contrastadas pela comparação dos grupos. Em cada um destes dois grupos, é verificado o número de indivíduos que foi exposto ao fator de risco, ou seja, ao fator de exposição, coletado geralmente de forma retrospectiva. É muito importante que o parâmetro que será estudado, tanto no grupo controle quanto no grupo de casos, seja avaliado da mesma forma. Este tipo de delineamento é recomendado quando o desfecho que se pretende estudar é raro para estabelecimento da etiologia quando o estudo de coorte se torna irrealizável.

Por exemplo, um estudo de caso-controle de mulheres no climatério com IAM (casos) e um grupo de mulheres saudáveis (controles), são avaliadas quanto ao uso prévio de TRH (Figura 10).

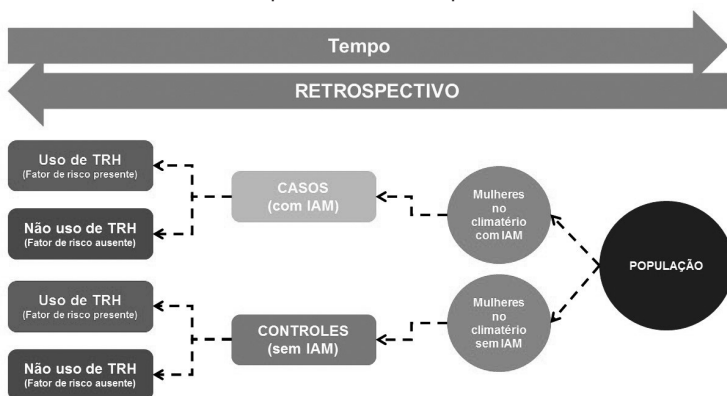


Figura 10. Exemplo de estudo de caso-controle retrospectivo.

Fases do estudo caso-controle

As fases principais de um estudo de caso-controle são as seguintes:

1. seleção de casos: identificação de pacientes com determinada doença ou condição de saúde, com base em um critério diagnóstico definido e acurado. Os casos são selecionados com base na doença, mas não na exposição;
2. pareamento com controles: pacientes com alguma outra doença, a população geral, vizinhos ou parentes. Os controles devem representar a prevalência da exposição na população onde os casos foram originados. Importante: os participantes da pesquisa do grupo controle devem ser selecionados do mesmo contexto em que foram selecionados os casos, como por exemplo, mesmo bairro, mesmo hospital ou região geográfica. A escolha de casos e controles não deve ser influenciada pelo nível de exposição, que deve ser determinada da mesma maneira para ambos. Quanto maior o número de casos, maior é o poder do estudo, pois será possível limitar melhor os fatores de confusão;
3. após determinação dos grupos de caso e de controles, é feita a coleta dos dados da exposição que se quer avaliar.

Seleção de casos e controles

Quem é considerado um caso? E, quem é considerado um controle? O primeiro passo para a realização de um estudo de caso-controle bem-sucedido é a seleção cuidadosa dos casos e dos controles.

O local de obtenção de casos e controles depende das características da doença em estudo. Um potencial viés do estudo de caso-controle é a definição incorreta de quem é considerado um caso e quem é considerado um controle. Deve-se definir um critério diagnóstico adequado para os casos, de forma a evitar que um caso seja classificado indevidamente como controle ou que um indivíduo sem a doença seja classificado como caso. Da mesma forma, quando escolhermos os controles devemos validar que ele é comparável aos casos, porém sem a condição de saúde de interesse.

A definição específica da doença é muito importante para garantir que a doença em questão está realmente presente entre os indivíduos que estão sendo definidos como casos. Para isto, precisamos estabelecer os critérios diagnósticos precisos e qual é o espectro da doença a ser estudada, considerando estabelecer a gravidade do caso. Podemos optar por excluir as formas mais leves da doença, ou ainda, melhorar a precisão do método diagnóstico que estamos utilizando, a fim de incluir estes indivíduos. Mas, indivíduos que têm uma forma mais avançada da doença são mais fáceis de serem diagnosticados, o que traz maior segurança ao estudo.

Ao selecionar os casos, existem alguns aspectos que devem ser considerados:

1. estabelecer o critério diagnóstico para a inclusão dos doentes no grupo de casos;
2. definir a origem dos pacientes incluídos no estudo (ambulatório, enfermaria, serviços médicos hospitalares diferentes, etc.);
3. definir os estágios da doença e registrá-los. Se houver variantes clínicas, caracterizá-las e registrá-las;
4. definir o procedimento de recrutamento dos doentes;
5. definir as condições e doenças concomitantes que indicam exclusão de pacientes. Por exemplo, tabagismo, etilismo, desnutrição, câncer, doenças sistêmicas, uso de medicamentos;
6. considerar grupos étnicos, sexo e faixa etária; e
7. estabelecer critérios que permitam agrupar os doentes em diferentes estratos.

Como o objetivo dos estudos de caso-controle normalmente é avaliar o desenvolvimento da doença, deve optar preferencialmente por casos incidentes (casos novos) da doença, ao invés de escolher casos mais prevalentes (casos crônicos), devido a dificuldade de determinar se a característica em estudo está relacionada ao prognóstico ou à sua causalidade ou a sobrevida desses indivíduos. A principal razão para preferir casos incidentes é estabelecer se a exposição de interesse esteve claramente presente antes da ocorrência da doença.

Em relação a constituição do grupo controle, temos de pensar que o controle deverá ter características similares ao grupo de casos, exceto pela presença da doença. Dessa forma, é indicado que:

1. o grupo controle proceda da mesma área geográfica e que apresente características socioeconômicas e culturais idênticas às do grupo de casos;
2. o grupo controle seja constituído por pessoas atendidas pelos mesmo serviço de origem dos casos;
3. caso não se opte pelo mesmo serviço, o grupo controle pode ser constituído por vizinhos, parentes ou colegas de trabalho ou de escola dos casos que estão sendo estudados;
4. quando o pesquisador dispõe de número suficiente de indivíduos controles, pode-se usar todos os indivíduos ou utilizar amostra do conjunto desses indivíduos (pareamento).

A interpretação dos resultados de um estudo caso-controle depende do pressuposto de que o grupo controle foi obtido de uma população adequada para estimar a frequência da exposição. Os controles devem ter a mesma oportunidade de serem selecionados como casos se, ao invés de saudáveis, fossem ou se tornassem doentes.

Um método para assegurar que os casos e os controles derivam da mesma população subjacente e que vão ter a mesma oportunidade de serem diagnosticados com uma doença é fazer um estudo caso-controle aninhado. Este desenho de estudo seleciona os casos e controles a partir de um estudo de coorte maior.

Caso-controle de base populacional aninhado

Os estudos de caso-controle de base populacional são aqueles em que os casos e controles são selecionados da população. Os casos podem ser detectados através de triagem populacional, em área geográfica definida durante um determinado período de tempo, ou pode-se utilizar registros médicos para se identificar todos os possíveis casos da área de estudo ou uma amostra

aleatória deles. Os controles são selecionados através de uma amostra probabilística dos indivíduos sem a doença, pertencentes à mesma área geográfica dos casos.

No estudo de caso-controle aninhado, casos e controles são selecionados no decorrer de uma coorte pré-definida, na qual algumas informações sobre exposições e fatores de risco já se encontram disponíveis. Utiliza-se como casos os indivíduos que se tornaram doentes no período, e como controles, uma amostra aleatória da coorte original, ou entre aqueles na coorte original, sem a doença. Para cada caso, controles são selecionados aleatoriamente de indivíduos que se encontram sob risco no momento do diagnóstico do caso, o que significa um emparelhamento pelo efeito de confusão do tempo. Informações adicionais são coletadas e analisadas no momento da seleção de casos incidentes e controles.

Pareamento ou emparelhamento dos grupos

O emparelhamento ou pareamento é uma técnica utilizada para controle de variáveis de confusão. A principal finalidade é garantir uma maior comparabilidade entre os grupos de casos e controles ou um maior controle dos fatores comuns que não são facilmente identificáveis, como sexo, idade, etnia, nível socioeconômico, etc. Em um estudo caso-controle, a doença de interesse é geralmente rara, logo, encontrar casos é muitas vezes o passo limitante do processo.

Não há regras específicas sobre o número de controles que são necessários em cada estudo. Para cada caso, podem ser escolhidos um, dois, três, controles. As principais variáveis de pareamento são a idade e o sexo, que são variáveis que geralmente estão associadas a possibilidade de exposição e desenvolvimento de uma doença, tornando os grupos mais comparáveis e minimizando potenciais distorções dos resultados na avaliação de riscos. Por este motivo, casos e controles são usualmente selecionados dentro da mesma faixa etária e sexo.

Quanto mais controles há no estudo, mais precisa a estimativa fornecida da frequência de exposição avaliada e também maior poder do estudo. O poder do estudo aumenta com até três

ou quatro controles para cada caso. A partir deste valor, a adição de mais algum controle acaba tendo pouco efeito sobre o poder do estudo. Um fator limitante do número de controles é o custo.

Amostragem

Deve-se verificar em que extensão o grupo escolhido representa todos os indivíduos com a doença, não só para interpretação dos resultados para os indivíduos incluídos no estudo (validade interna), como também para se poder fazer inferências, extrapolações para indivíduos não incluídos no estudo (validade externa).

O que define o tamanho da amostra serão o número de casos e controles que serão selecionados para testar a hipótese de estudo. O tamanho da amostra é inversamente proporcional ao risco que se pretende detectar: quanto maior o risco da condição em estudo, menor o tamanho da amostra, e aqueles estudos que tem poucos casos irão apresentar baixo poder estatístico para detectar riscos.

Para o cálculo do tamanho da amostra é necessário definir:

1. nível de significância (α) do teste, geralmente de 5%;
2. poder do teste ($1-\beta$), geralmente de 80-90%;
3. proporção de pessoas expostas ao fator de risco na população geral;
4. valor de *Odds Ratio* (OR) mínimo a ser detectado;
5. razão entre o número de controles e casos.

Para avaliar as medidas de exposição pode-se utilizar entrevistas, questionários, informações de parentes/vizinhos, marcadores biológicos, exames laboratoriais, etc. Entretanto, o importante é que a forma de obtenção desta informação seja da mesma forma para os grupos caso e controle e o mascaramento ou cegamento do pesquisador, minimizando o viés do observador. Também é importante determinar o início e a duração da exposição, tanto para os casos quanto para os controles. No delineamento de estudos de caso-controle, o nível de exposição de casos é usualmente determinado após o desenvolvimento da doença

(dados retrospectivos) e, em geral, pelo questionamento direto à pessoa afetada ou a um parente ou amigo (viés de memória).

Medida de associação: *Odds Ratio*

A medida de associação utilizada no estudo de caso-controle é o *Odds Ratio* (OR), que é a razão do *odds* de exposição entre os casos dividido pelo *odds* de exposição entre os controles. Isto é, a razão de probabilidade entre estes dois quocientes, complementares entre si. O OR é utilizado como uma aproximação de razão de riscos (RR), quando as incidências são baixas, ou quando a aproximação da razão de prevalência (RP), quando as prevalências são baixas também.

Quando os casos e controles são selecionados da população em geral, a proporção dos expostos ao fator de risco no grupo controle pode ser utilizada como uma estimativa da proporção de expostos na população em geral (RAP). Ou seja, não podemos calcular o RAP, mas podemos calcular uma estimativa deste risco atribuído. Lembre-se que o RAP nos informa a proporção da doença na população de estudo, que é atribuível à exposição ou fator de exposição, e que poderia ser eliminada se o fator fosse removido.

O resultado amostral de um estudo de caso-controle pode ser expresso em uma tabela de contingência 2 x 2 (Figura 11).

		DESFECHO (doença)	
		CASO (doente)	CONTROLE (não doente)
FATOR	EXPOSTO (presente)	a	b
	NÃO EXPOSTO (ausente)	c	d

Figura 11. Tabela de contingência 2 x 2 do estudo de caso-controle.

Na tabela, temos (Figura 12):

- o *odds* da presença de exposição ao fator, entre os casos, que é: a / c ;
- o *odds* da presença de exposição ao fator, entre os controles, que é: b / d .

$$\text{Odds Ratio (OR)} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c}$$

Figura 12. Cálculo do OR.

Quando a doença é rara, a razão de chance de doença (OR) e o RR são semelhantes. A interpretação do OR é semelhante ao RR:

- OR = 1,0: indica que probabilidade de doença nos expostos ao fator de risco e nos não expostos são equivalentes;
- OR > 1,0: indica que a exposição ao fator em estudo é de risco, podendo implicar em relação causa-efeito;
- OR < 1,0: indica que a exposição ao fator em estudo é de proteção.

Vantagens e desvantagens dos estudos de caso-controle

Os estudos de caso-controle têm grande aplicação em situações em que a doença é relativamente pouco frequente (rara ou de baixa incidência), difíceis de serem estudadas em um estudo de coorte, e o tempo decorrido entre a exposição ao risco e a evidenciação do seu efeito (desfecho) é longo. É um estudo rápido, prático e de baixo custo para testar o efeito de fatores que se relacionam com o evento estudado. Os estudos caso-controle têm limitada implicação ética, uma vez que não existe intervenção nem observação prospectiva de exposições de risco.

Os estudos de coorte identificam indivíduos com base em seu *status* de exposição e, subsequentemente, seguem os indivíduos para observar o desfecho de interesse. Em comparação, os estudos de caso-controle identificam indivíduos com base em seu *status* de doença, permitindo o estudo de várias exposições (múltiplas exposições) dentro de um grupo pré-definido de

casos e de controles. São úteis quando dados anteriores sobre a exposição estão disponíveis ou podem ser facilmente obtidos e, ao contrário dos estudos de coorte, pelo fato de não necessitar de um grande número de participantes, pode-se empregar exames e testes mais dispendiosos e laboriosos. Comparado ao estudo de coorte, também tem menor perda de seguimento de casos.

As principais limitações dos estudos de caso-controle são: delineamento observacional (fatores de confusão); não são adequados para investigações de exposições raras, a menos que o RAP seja muito alto; não estima a incidência das doenças e; a informação sobre a exposição ou fator é obtida após a ocorrência da doença e, portanto, não há como se distinguir uma cronologia nítida entre a exposição e o aparecimento da doença, diferente do que ocorre nos estudos de coorte onde a exposição é o ponto de partida.

Vieses dos estudos de casos-controle

A classificação de indivíduos como casos, por terem a doença de estudo, e como controles, por não terem a doença, deve seguir critérios bem definidos, a fim de evitar viés de classificação. Este é um erro sistemático pelo qual doentes são selecionados como controles e indivíduos sem a doença são selecionados como caso. Exames laboratoriais, altamente sensíveis e específicos, são desejáveis, em complementação a diagnósticos clínicos. Este procedimento visa a minimizar a classificação de não doentes como casos (falsos casos) e a de doentes como controles (falsos controles).

O viés de seleção é uma grande preocupação em estudos caso-controle, quando a seleção de casos e de controles é baseada em critérios diferentes e estes estão relacionados à exposição, afetando a comparabilidade de casos e controles. Um dos motivos deve-se ao fato de que os princípios da comparabilidade na seleção de controles geralmente são influenciados pelo princípio da eficiência, ou seja, disponibilidade de recursos e tempo.

Outro problema inerente aos estudos de caso-controle é o viés devido à natureza retrospectiva da informação do fator de risco. Por exemplo, se as informações sobre exposição provêm de entrevistas, os casos geralmente relatam tais informações depois

de conhecer seu diagnóstico. Ou seja, o diagnóstico pode afetar o relato de várias maneiras, por exemplo, melhorar a memória ou provocar falsas memórias de exposição (viés de recordação ou de memória). Uma forma de evitar o viés de recordação é extrair as informações de registros criados antes da ocorrência da doença.

Ao selecionar os controles deve-se assegurar que as observações sejam realizadas nos dois grupos sob as mesmas condições. O ideal é que o pesquisador não tenha conhecimento (mascaramento) de quem tem a doença e quem não tem, ou seja, quem é caso e quem é controle, para evitar tendenciosidades na coleta de informações, ou seja, o viés do observador.

A seleção de casos prevalentes ao invés de casos incidentes também é outra fonte potencial de viés. A prevalência é afetada pela duração da doença que é influenciada pelo tratamento e cura e também pela mortalidade associada à doença. Quando se inclui casos prevalentes, os fatores podem estar estatisticamente associados à doença devido a um "efeito de sobrevivência" e duração da doença e não devido à uma associação causal.

Redação dos estudos de caso-controle

Como mencionado, o guia STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*) (www.equator-network.org/reporting-guidelines/strobe), disponível na base EQUATOR, traz um *checklist* de verificação para a redação de estudos de coorte, caso-controle e transversal.

Análise crítica dos estudos de caso-controle

As principais questões para avaliarmos um estudo de caso-controle são:

1. Como foi realizada a seleção dos casos?

a. qual o critério diagnóstico estabelecido para a inclusão dos doentes no grupo de casos? O processo mais propenso ao viés não é a avaliação do desfecho, mas o diagnóstico de "caso típico" e a decisão de quando o indivíduo se tornou um caso.

b. qual a origem dos casos incluídos no estudo? Casos podem ser identificados em hospitais, ambulatórios (geral

ou de especialidades), clínicas especializadas ou serviços de saúde (diferentes).

c. qual foi o procedimento de recrutamento dos doentes? Os pesquisadores devem definir explicitamente os critérios para o diagnóstico de um caso e os critérios de elegibilidade utilizados para a seleção.

d. há condições e doenças concomitantes que indicam exclusão de pacientes? Fatores como tabagismo, etilismo, desnutrição, cânceres, doenças sistêmicas, uso de medicamentos, etc., podem levar a exclusão de pacientes.

2. Como foram definidos os controles?

a. qual a origem do grupo controle incluído no estudo? A interpretação dos resultados de um estudo de caso-controle depende do pressuposto de que o grupo controle foi obtido de uma população adequada para estimar a frequência da exposição. Os controles devem vir da mesma população que os casos e sua seleção deve ser independente das exposições de interesse. Questione se o controle presente fosse um caso, este seria encontrado onde os casos estão sendo encontrados? Controles podem ser recrutados nos hospitais onde os casos foram obtidos, na vizinhança dos casos, nas mesmas escolas, entre amigos e colegas de trabalho dos casos, na população em geral.

b. foi realizado pareamento? Os estudos de caso-controle usam o pareamento para aumentar o grau de semelhança entre os casos e os controles. O pareamento pode reduzir a possibilidade de que outros fatores atrapalhem a associação entre a exposição e o desfecho (fator de confusão).

3. Os resultados obtidos pelo estudo são relevantes?

a. os potenciais fatores de confusão foram identificados e controlados na análise dos dados? Os casos podem diferir dos controles em outros fatores além da exposição de interesse. O confundimento ocorre quando um outro fator, que não a exposição de interesse, distorce a associação entre a exposição e o desfecho, limitando assim a inferência de que a exposição causa a doença. Os pesquisadores devem definir os fatores de confusão em estudos de caso-

controle na fase de concepção (pareamento) ou, se não for possível ou desejável o pareamento, os fatores devem ser ajustados com técnicas analíticas (análise multivariada).

b. houve cegamento dos pesquisadores na coleta dos dados de exposição em relação ao desfecho? Os pesquisadores que forem coletar os dados devem estar mascarados (cegados) para o *status* de caso ou controle dos participantes ou, se não possível, mascarados para a hipótese principal do estudo. Os pesquisadores precisam de treinamento para aferir a exposição de um modo semelhante a partir de casos e controles (padronização).

c. o estudo de caso-controle teve amostra adequada? Há cálculo de tamanho de amostra? Estudos conduzidos com poucos casos apresentam baixo poder estatístico para detectar riscos.

d. a análise estatística realizada foi adequada? Estudos de caso-controle fornecem uma estimativa (aproximação) sobre o RR (razão de chance ou OR) da doença. Os estudos de caso-controle podem estimar os RAP a uma doença, mas não podem determinar a incidência absoluta da doença nem o risco atribuível.

4. os dados podem ser generalizados? Atenção meticulosa deve ser dada à esses pontos para garantir a validade dos resultados e reforçar a confiança das conclusões do estudo.

Estudo transversal

O estudo transversal é indicado para responder perguntas como: qual a prevalência da doença em estudo? Quais são as frequências do fator de risco ou de exposição em estudo? Existe associação entre o fator de risco ou de exposição e o desfecho em questão?

Também conhecido como estudo seccional, o estudo transversal tem como objetivo avaliar um indivíduo em um determinado momento no tempo, ou seja, o estudo não apresenta temporalidade. No entanto, pode-se referir a coleta de dados de forma retrospectiva, quando os dados são coletados de prontuários ou registros, ou prospectiva, quando se está planejando o estudo

e pretende-se aplicar um questionário ou inquérito, de modo a situar o leitor quanto à obtenção das informações.

O estudo transversal é um estudo primário, podendo ser analítico ou descritivo, controlado ou não controlado (indivíduos não expostos) e é exclusivamente observacional (não intervencional), realizado em um único momento. Ou seja, as medidas de exposição e de efeito são realizadas ao mesmo tempo e individualmente. Tipos especiais de estudos transversais incluem estudos de prevalência, de inquérito, de teste diagnóstico ou de rastreamento e ecológicos.

Os estudos de prevalência estimam a prevalência (frequência) de uma determinada característica (doença ou condição) de uma população de interesse e fatores associados ao desfecho. São pesquisas por amostragem, nas quais o objetivo principal é a estimação, e é fundamental uma amostra representativa da população. Apresentam pouca ou nenhuma perda, gerando resultados que não são enviesados.

Os estudos de testes diagnósticos ou de rastreamento, como visto no capítulo 5, são utilizados para avaliar a performance de um novo teste diagnóstico ou de rastreamento (teste índice) comparado a um padrão de referência (padrão-ouro). Dois grupos de indivíduos são selecionados, com e sem doença de acordo com o padrão-ouro. É avaliada a acurácia (sensibilidade e especificidade) do novo método, que fornece um resultado mais rápido, mais barato ou menos evasivo do que o padrão-ouro. A sensibilidade e a especificidade são calculadas como a proporção de pacientes com doença cujos testes foram positivos e a proporção de pessoas sem doença com teste negativo, respectivamente. Lembre-se que é importante que haja cegamento do teste índice em relação ao resultado do padrão-ouro, ou seja, que ao se fazer a avaliação do novo teste diagnóstico, os pesquisadores que realizam o novo teste tenham desconhecimento dos resultados obtidos com o padrão-ouro.

Já no estudo ecológico, abordado em mais detalhes ao final deste capítulo, as medidas da exposição e da doença são observadas em grandes populações e não mais para cada indivíduo do estudo. A exposição é conhecida somente para os grupos, e não para os indivíduos.

O que é prevalência?

A prevalência envolve os casos existentes em um determinado momento, incluindo-se casos novos e casos antigos naquele "recorte", enquanto a incidência corresponde aos casos novos de uma determinada doença. A prevalência de uma doença ou condição de saúde é medida pelo cálculo da proporção entre o número de pessoas acometidas (casos) e as que não são acometidas.

População de estudo

A escolha e a definição da população de estudo depende da pergunta de pesquisa do pesquisador, das características da doença de interesse e da própria população que se deseja estudar. O desenho amostral deve garantir a representatividade da amostra de estudo (população-alvo).

Tamanho da amostra

Para o cálculo do tamanho da amostra são necessárias:

1. estimativa da proporção esperada da doença na população;
2. amplitude (margem) do intervalo de confiança que se assume como o erro máximo aceitável;
3. definição do intervalo de confiança, geralmente de 95%.

Fonte de dados

Pode-se obter tanto dados primários, como questionários, inquéritos e entrevistas, etc., quanto dados secundários, como registros médicos, prontuários eletrônicos, etc. Idealmente, a obtenção de toda informação deve ocorrer no mesmo momento. Por exemplo, pode-se definir que é interessante obter a informação da primeira consulta com o clínico. Entretanto, sabemos que a obtenção da informação completa necessária para condução do estudo dificilmente é coletada em uma única consulta, com o número de consultas podendo variar. É preciso ter cuidado para não coletar uma informação de datas em um período longo de tempo (por exemplo, com um intervalo de um ano), pois o objetivo do estudo é analisar apenas um ponto específico do tempo. Para

isso, pode-se determinar o intervalo de tempo da informação coletada (por exemplo, um desvio de um mês para mais e um mês para menos a partir do ponto de coleta de dados definido). Esta informação deve ser incluída na seção de métodos do estudo.

Fatores de exposição e desfecho

Fatores de exposição estão relacionados com o desfecho. Os fatores de exposição geralmente são variáveis independentes, e o(s) desfecho(s) corresponde(m) a(s) variável(is) dependente(s).

Modelo conceitual do estudo transversal

Um estudo transversal ocorre quando uma população ou amostra de sujeitos é estudada em um único instante no tempo, como por exemplo, no Censo do IBGE. Ou seja, em que a exposição e o desfecho estão presentes no mesmo momento (Figura 13). Dessa maneira, o estudo transversal caracteriza-se pelo fato de que a exposição e o desfecho estão ocorrendo simultaneamente, embora a exposição possa existir só no momento atual, ou existir desde algum tempo no passado ou, por fim, ser uma característica do indivíduo. Portanto, dificulta-se o conhecimento da relação temporal existente entre exposição e desfecho.

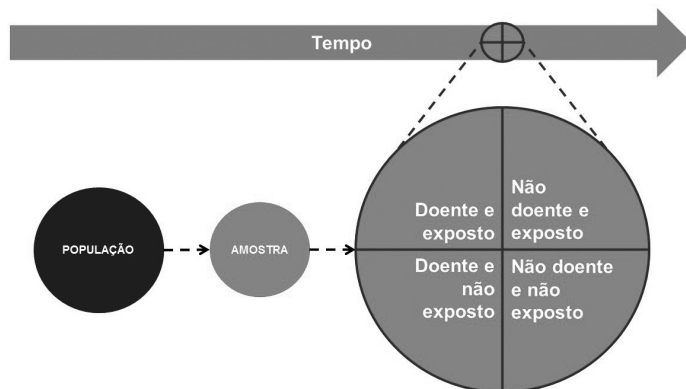


Figura 13. Modelo conceitual do estudo transversal.

Por exemplo, um estudo transversal de mulheres no climatério avaliadas em um momento, observa a prevalência de IAM naquelas que receberam TRH e nas que não receberam TRH.

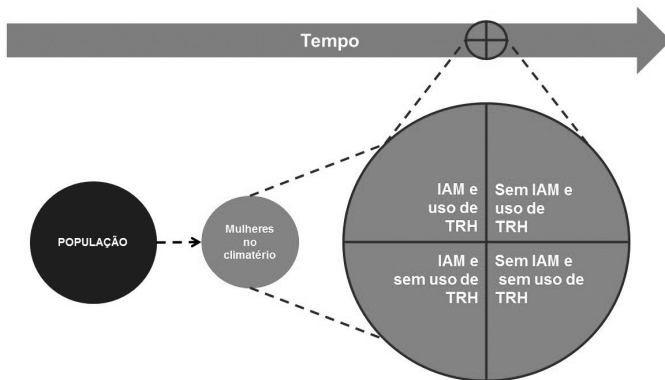


Figura 14. Exemplo de um estudo transversal.

Fases do estudo transversal

As fases principais de um estudo transversal podem ser resumidas em:

1. definição de uma população de interesse;
2. estudo da população por meio da realização de amostragem;
3. determinação da presença ou ausência do desfecho e da exposição para cada um dos indivíduos estudados em um único momento.

Razão de prevalência

Considerando a exposição e o desfecho como sendo variáveis dicotômicas, pode-se construir e apresentar os elementos em uma tabela de contingência 2 x 2 (Figura 15), onde teremos:

- indivíduo exposto, que apresenta o desfecho;
- indivíduo exposto, que não apresenta o desfecho;
- indivíduo não exposto, que apresenta o desfecho e;
- indivíduo não exposto, que não apresenta o desfecho.



Figura 15. Tabela de contingência 2 x 2 do estudo transversal.

Se desejarmos estimar a prevalência do desfecho na população como um todo, basta proceder os cálculos da seguinte forma: $a + c / a + b + c + d$, onde "a" corresponde aos indivíduos com desfecho expostos, "c" aos indivíduos com desfecho não expostos e " $a + b + c + d$ " ao total da amostra de estudo.

$$\text{Prevalência global do desfecho} = \frac{a+c}{a+b+c+d}$$

Figura 16. Prevalência global do desfecho.

Também há a possibilidade de estimar a prevalência do desfecho entre expostos ao fator de risco ou de exposição: $a / a + b$, onde "a" corresponde aos indivíduos com desfecho expostos e " $a + b$ " ao total de indivíduos expostos. Esta medida indica a prevalência de portadores do desfecho entre todos os indivíduos expostos do estudo.

$$\text{Prevalência do desfecho entre expostos} = \frac{a}{a+b}$$

Figura 17. Prevalência do desfecho entre expostos.

E a prevalência do desfecho entre não expostos ao fator de risco: $c / c + d$, em que "c" indica os indivíduos com desfecho não expostos e " $c + d$ " o total de indivíduos não expostos. Este cálculo fornece a prevalência do desfecho entre os não expostos do estudo.

$$\text{Prevalência do desfecho entre não expostos} = \frac{c}{c+d}$$

Figura 18. Prevalência do desfecho entre não expostos.

Ao dividirmos uma fração pela outra, teremos: $(a / a + b) / (c / c + d)$, que indica a razão de prevalência (RP). No caso, se o numerador da fração for maior do que o denominador, isto constituirá evidência de que a frequência do desfecho foi maior entre os expostos e que há uma possível associação entre ambos. Uma maior prevalência do desfecho estará associada com a presença da exposição em estudo.

$$\text{Razão de prevalência} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Figura 19. Razão de prevalência.

A interpretação é a mesma do RR e OR:

- RP = 1,0: indica que não há associação. As prevalências da doença nos grupos de expostos e não expostos são idênticas, indicando que não há associação observada entre exposição e doença (hipótese nula);
- RP > 1,0: indica que há associação positiva entre o grupo de expostos ao fator estudado e o desfecho, em relação ao grupo de não expostos;
- RP < 1,0: indica que há associação negativa (inversa) entre o grupo de expostos ao fator estudado e o desfecho, em relação ao grupo de não expostos.

Existem fatores que aumentam a prevalência, como a maior duração da doença, aumento da incidência, aumento das taxas de sobrevivência de uma doença (sem cura), melhoria dos métodos diagnósticos, melhoria dos sistemas de informação, imigração de casos de uma determinada doença e, emigração de indivíduos sadios. Por outro lado, há fatores que diminuem a prevalência, como migração de pessoas sadias, emigração de casos de doença, diminuição das taxas de incidência, maior letalidade, aumento das taxas de cura de uma determinada doença e, menor duração da doença.

Quais são os critérios para julgar causalidade?

A presença de uma associação estatística não significa necessariamente uma associação causal. Existem critérios utilizados para julgamento da causalidade, definidos por Hill (1965):

- força da associação: quanto mais forte uma associação, maior será a possibilidade de se tratar de uma relação causal;
- temporalidade: a causa deve sempre preceder o efeito;
- consistência ou replicação: mesmo resultado obtido em diferentes circunstâncias. Relação condizente com achados de outros estudos, em diferentes tipos/desenhos de estudo, populações e circunstâncias;
- gradiente biológico: efeito dose-resposta.

Vantagens e desvantagens do estudo transversal

São estudos de baixo custo, de fácil (simplicidade) realização, rapidez na execução e objetividade na coleta de dados. Facilmente se obtém amostras representativas da população e são úteis na investigação das exposições que são características individuais fixas. Na investigação de surtos epidêmicos, a realização de um estudo transversal medindo diversas exposições é, em geral, o primeiro passo para a determinação da sua causa.

Uma limitação importante do estudo transversal é que ele informa pouco sobre a ordem dos eventos (desfecho ou exposição). Ou seja, devido ao fato de a exposição e o desfecho serem coletados em um único momento no tempo, torna-se mais difícil estabelecer uma relação temporal entre os eventos e considerar com maior grau de certeza se a relação entre eles é causal ou não. Dificuldade para investigar condições de baixa prevalência, pois exige amostra grande e indivíduos que possuem o desfecho por um maior período de tempo têm maior chance de serem incluídos na pesquisa. Ou seja, os casos de maior duração são mais facilmente recrutados para investigação e os agravos à saúde de curta duração têm menos chance de aparecerem nos resultados, tornando difícil diferenciar se as exposições estudadas estão associadas ao surgimento de novos casos ou à duração dos mesmos (viés de prevalência).

Por se tratar de um estudo observacional, torna-se impossível assegurar que os fatores de confusão vão estar igualmente distribuídos entre os grupos. Os grupos (expostos doentes, expostos não doentes, não expostos doentes e não expostos não doentes) podem terminar tendo tamanhos muito diferentes, resultando em perda da eficiência estatística.

Vieses do estudo transversal

O viés de seleção é o principal problema do estudo transversal. O viés de sobrevivência seletiva ocorre quando casos prevalentes são usados para estudar a associação exposição-doença. Casos prevalentes representam sobreviventes da doença em questão e, como sobreviventes, podem ser atípicos. O viés de informação também é comum em estudos transversais.

O viés de aferição ou de mensuração ocorre por variação entre observadores; inadequação do entrevistador ou do informante; deficiência ou variação dos instrumentos de mensuração ou coleta de dados; erros técnicos de aferição; forma de detecção de diagnóstico; processamento da informação (codificação/digitação). O viés de memória ocorre quando a história de exposição é obtida retrospectivamente e os casos podem lembrar-se melhor de sua história de exposição. Depende também do intervalo de tempo entre a exposição e o desfecho.

Redação de estudo transversal

O guia STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*) (www.equator-network.org/reporting-guidelines/strobe), traz um *checklist* de verificação para a redação de estudos de coorte, caso-controle e transversal.

Análise crítica do estudo transversal

As principais questões norteadoras para avaliar um estudo transversal são:

1. qual(is) o(s) desfecho(s) e fator(es) de exposição? Como foram mensurados? Quais foram as variáveis dependentes e independentes do estudo? O estudo deve descrever de que forma foram mensurados o(s) desfecho(s) e fator(es)

de exposição (questionário, prontuário eletrônico, marcadores biológicos, etc.);

2. como a amostra foi recrutada? Houve representatividade da amostra estudada? Foi realizado cálculo amostral? A amostra deve garantir a representatividade da população-alvo e deve considerar a pergunta de pesquisa, as características da doença e da população;

3. a análise dos dados foi adequada? O estudo deve mencionar as medidas de frequência (RP) e de associação;

4. houve associação entre exposição e desfecho? Os autores afirmaram que essa associação é causal? Pode-se medir a associação entre exposição e desfecho, mas há limitação para inferir causalidade pois não há temporalidade, um critério importante para se estabelecer causa. Portanto, é inapropriado estabelecer relação causal em estudo transversal;

5. os dados podem ser generalizados? Para se generalizar, é necessário assumir que a amostra de estudo é idêntica à população.

Estudo Ecológico

Também chamados de estudos de correlação, o estudo ecológico é um tipo de estudo transversal. A diferença é que avaliamos um grupo de indivíduos em um determinado momento no tempo, então a unidade de análise deixa de ser o indivíduo e passa a ser um grupo de indivíduos.

Os estudos ecológicos fornecem evidências mais fracas sobre causalidade em virtude de serem feitas extrapolações incorretas para os indivíduos a partir de dados regionais ou nacionais. Para exposições que normalmente não podem ser medidas individualmente (como poluição do ar, resíduos de pesticidas nos alimentos, fluoretação da água) os estudos ecológicos são muito importantes. Quando a relação causal já foi bem estabelecida, estudos ecológicos bem delineados podem ser muito úteis.

Estudos ecológicos também podem ser feitos comparando-se populações em diferentes lugares ao mesmo tempo ou, em uma série temporal, comparando-se a mesma população em diferentes momentos. Os estudos de série temporal podem reduzir o efeito de confusão causado pelo nível socioeconômico, que é um potencial problema nos estudos ecológicos. Se o período de tempo em um estudo de série temporal for muito curto, como em um estudo de série temporal diária, o fator de confusão é praticamente inexistente, com os participantes do estudo servindo como seus próprios controles.

Embora fáceis de realizar, os estudos ecológicos são frequentemente difíceis de interpretar, uma vez que raramente é possível encontrar explicações para os resultados obtidos. Em geral, os estudos ecológicos baseiam-se em dados coletados com outros propósitos (dados secundários). Desta forma, os dados obtidos de diferentes exposições e de fatores socioeconômicos podem não estar disponíveis. Além disso, uma vez que a unidade de análise é uma população ou um grupo populacional, a relação entre exposição e efeito no nível individual não pode ser estabelecida.

Um atrativo dos estudos ecológicos é que podem ser utilizados dados de diferentes populações com características muito diferentes ou extraídos de diversas fontes de dados. Este tipo de estudo é particularmente propenso a fatores de confusão, e também a relação entre o fator de exposição e o evento pode não estar ocorrendo ao nível do indivíduo. Por isso, a associação entre uma exposição e uma doença, ao nível da população, não permite afirmar que a exposição está mais presente naqueles que adquirem a doença (falácia ecológica). Entretanto, os estudos ecológicos ajudam a identificar fatores que merecem uma investigação detalhada.

A falácia ecológica ou viés ocorre quando são tiradas conclusões impróprias com base em estudos ecológicos. O viés ocorre porque a associação observada entre as variáveis no nível de grupo não representa, necessariamente, a associação existente no nível individual.

Referências

Bastos & Duquia. Um dos delineamentos mais empregados em epidemiologia: estudo transversal. *Scientia Medica*. 2007;17(4):229-232.

Bonita, R.; Beaglehoe, R.; Kjellström, T. *Epidemiologia básica*. 2. ed. São Paulo: Santos, 2010-2019. 213 p.

Greenhalgh, T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Hill, B.A. The Environment and Disease: Association or Causation? *Proc R Soc Med*. 1965;58(5): 295–300.

Oliveira, M.A.; Vellarde, G.C.; Moreira de Sá, R.A. Entendendo a pesquisa clínica IV: estudos de caso controle. *FEMINA*. 2015;43(4):175-180.

Oliveira, M.L.; Vellarde, M.C.; Moreira de Sá, R.A. Entendendo a pesquisa clínica III: estudos de coorte. *FEMINA*. 2015;43(3):105-110.

Pinho de Oliveira & Parent. Estudos de Coorte e de Caso-Controlle na Era da Medicina Baseada em Evidência. *Bras. J. Video-Sur*. 2010;3(3):115-125.

Estudos descritivos (relato e série de casos) e estudos pré-clínicos (*in silico*, *in vitro* e *in vivo*)

*Sinara Santos
Vânia Marisia Santos Fortes dos Reis
Edison Capp
Otto Henrique Nienov*

Relato de caso e série de casos

O relato de caso e a série de casos foram durante muito tempo a única base de informações científicas disponível. Com o surgimento da medicina baseada em evidências, passaram a ser considerados de baixo nível de evidência, pois limitam-se a descrição detalhada do quadro clínico e das características, geralmente de doenças raras, de um ou alguns indivíduos acometidos por uma doença (casos).

Desta forma, o relato de caso e a série de casos correspondem a um estudo de delineamento puramente descritivo, sem grupo controle, de caráter narrativo e reflexivo, cujos dados são provenientes da prática cotidiana ou da atividade profissional. Ou seja, no momento da elaboração do relato de caso, os eventos narrados estarão findados, não estando previstos experimentos como objeto do estudo. Tem como finalidade destacar algum fato inusitado e relevante ou evento clínico raro ou uma nova terapia ou intervenção, ampliando tanto o conhecimento como sugerindo hipóteses para outros estudos.

Geralmente, são a primeira fonte de evidências para novas terapias (cirúrgicas ou clínicas) e para detecção de efeitos adversos raros de medicamentos. A razão da publicação pode corresponder à raridade do caso, a alguma peculiaridade, ou a algum aspecto terapêutico de interesse.

Qual a diferença entre relato de caso e série de casos?

O relato de caso corresponde a descrição detalhada de aspectos de interesse de um único paciente. A descrição pode conter dados da história de saúde, exame físico, sinais, sintomas, evolução, resultados de exames complementares e outras características do paciente, relatando os procedimentos terapêuticos utilizados, bem como o desfecho do caso. Possuem indicação em situações de doenças raras, em que tanto o diagnóstico como a terapêutica não estão claramente estabelecidos na literatura científica, assim como para situações peculiares ou aspecto terapêutico de interesse.

Freqüentemente, os relatos de casos são reunidos para formar uma série de casos, em que são descritos mais de um paciente com uma condição específica para ilustrar um aspecto do problema, do tratamento ou, o que é mais comum, das reações adversas ao tratamento. Ou seja, a série de casos agrega casos individuais em um relato único. O aparecimento de vários casos semelhantes em um curto período de tempo pode anunciar uma epidemia. O relato de apenas um caso incomum pode não desencadear uma investigação mais aprofundada, diferente da série de casos. Uma série de casos pode constituir um grupo caso de um estudo de caso-controle, que pode então permitir um maior conhecimento da doença.

Embora a diferenciação entre os dois tipos de estudos seja subjetiva e divergente entre diversos autores, há quem indique que um relato de caso engloba não mais do que três casos e uma série de casos compreende de três a dez ou mais casos. Por outro lado, alguns autores denotam que o relato de caso compreende até nove casos e de que a série de casos é um estudo com maior número de participantes (acima de dez) e pode ser retrospectivo ou prospectivo.

Indicação para elaboração do relato de caso e da série de casos

A principal pergunta que deve ser feita por quem cogita relatar um caso ou uma série de casos é: o relato contribui de forma substancial para a compreensão e o tratamento dessa doença ou de uma nova doença?

Como indicações claras para elaboração de relato de caso e série de casos teríamos: detecção de epidemias, descrição de características de novas doenças, formulação de hipóteses sobre possíveis causas para doenças, descrição de resultados de terapias propostas para doenças raras e de efeitos adversos raros em doenças comuns.

Um bom relato de caso deve ter o objetivo de acrescentar benefícios às práticas atuais ou de esboçar possíveis novas direções na pesquisa de determinado tema em que um único ou poucos indivíduos possam ser representativos. O relato deve propor formas inovadoras na abordagem de uma doença ou tratamento, além de formular novas hipóteses que podem ser testadas em outros desenhos de estudo.

Estrutura do relato de caso e da série de casos

Apesar de ser um estudo puramente descritivo, um relato de caso ou uma série de casos não são dispensados de trazer uma pergunta de pesquisa relevante, clara, definida, interessante e nova. Ou seja, apresentam um objetivo definido. Por conseguinte, existe um padrão de apresentação para um relato de caso ou uma série de caso, exceto que não possuem um método descrito.

Um relato de caso ou uma série de casos descrevem a história médica de um único paciente ou de um grupo de pacientes em forma de uma história. Como toda publicação, o texto deve ser simples, claro, preciso e conciso. A estrutura básica do relato de caso inclui título, resumo, palavras-chave, introdução com objetivo, descrição do caso, técnica, terapia ou situação, discussão com revisão da literatura, conclusão e bibliografia. Adicionalmente, pode-se fazer uso de figuras, tabelas, gráficos e ilustrações que complementam a descrição do caso. Frequentemente, o manuscrito deve ter cerca de 1.500 a 2.500 palavras e no máximo 20 a 30 referências.

O título deve ser sucinto e descritivo. O resumo deve conter entre 150 e 200 palavras, e contemplar todos os itens descritos no texto. A introdução deve ser concisa e apresentar o contexto, informação disponível sobre o tema e justificar porque o relato é relevante, além do objetivo. As frases e parágrafos devem ser nexos

e completos, sem informações truncadas ou soltas. O caso deve ser descrito com todos os detalhes relevantes e de forma sucinta.

Na exposição do caso, a sequência deve ser cronológica, organizada, com detalhes suficientes para compreensão do caso e eliminando dados demais, como detalhes de datas dos exames, dados confusos ou não confirmados. No caso de uma série de casos, estes devem ser sequencialmente relatados. Qualquer indicação para a identificação do paciente deve ser suprimida. A descrição deve incluir dados demográficos (idade, massa corporal, sexo, cor, ocupação), história clínica, comorbidades, exame físico, exames complementares alterados e desfecho de interesse. Deve-se evitar incluir evolução diária, interconsultas e exames rotineiros normais. O autor deve estabelecer uma relação temporal e causal com a situação relatada. As datas de referência ou de ocorrência de eventos devem ser relativas à internação ou intervenção principal, evitando-se colocar como referência a data de nascimento ou data real do atendimento. Os laudos de exames, procedimentos e necropsia não devem ser transcritos na íntegra, mas apenas conter as informações mais importantes à interpretação do caso. Exames clínicos ou laboratoriais negativos só devem ser incluídos quando absolutamente necessários à compreensão do caso. A intervenção, caso tenha, deve ser descrita em detalhes permitindo ser reproduzida por outros pesquisadores. Se for uma medicação, devem ser descritos nome comercial, nome do laboratório, dosagem, esquema de administração e duração do tratamento. Os possíveis efeitos adversos e interações também devem ser reportados. Em se tratando de técnica cirúrgica, esta deve também ser descrita minuciosamente para poder ser reproduzida.

A discussão deve enfatizar a prioridade e singularidade do relato, a acurácia do diagnóstico e a sua validade em comparação com os dados da literatura, e os subsídios para se levantar novas perspectivas, aplicações ou conhecimentos com o mesmo, além de potenciais confundidores. A conclusão do relato é limitada ao desenho de estudo, ou seja, deve ser compatível com as limitações de um relato de caso. As opiniões sobre a etiologia, prevenção e terapêutica, frequentemente encontradas nas conclusões destes artigos, devem ser vistas como o primeiro elo de uma cadeia de evidências a serem obtidas, e não como opinião definitiva.

Vantagens e desvantagens do relato de caso e da série de casos

Uma situação em que classicamente o relato de caso e a série de casos tem uma grande importância é no estudo de doenças raras. Isso se deve ao fato de ser praticamente impossível a compilação de vários casos de pacientes em um único estudo. Por muitas vezes, servem como alarme inicial para efeitos colaterais não vistos em estudos pré-clínicos e clínicos. Também terapêuticas são avaliadas quanto ao sucesso e ao fracasso num único indivíduo e esses “erros e acertos” são utilizados para um próximo caso. Neste caso, podem ser de boa qualidade se baseados em um conjunto de casos, onde há uniformidade de tratamento. São imediatamente compreensíveis e escritos e publicados rapidamente. Embora esse tipo de pesquisa seja tradicionalmente considerado uma evidência científica fraca, muitas informações que, de outro modo, seriam perdidas em um ensaio clínico ou em um levantamento, podem ser encontradas em um relato de caso.

Possuem limitações importantes e podem levar a conclusões equivocadas, uma vez que frequentemente avaliam dados retrospectivos, baseiam-se em poucos casos, não possuem grupos de comparação e apresentam resultados que se aplicam somente àquele paciente ou grupo de pacientes específicos. Não podem ser a base para avaliar etiologia e não quantificam a prevalência da doença na população. Somente podem demonstrar eficácia de um tratamento sob raríssimas condições. Jamais podem ser utilizados para demonstração de segurança de uma intervenção devido à raridade de alguns efeitos adversos e não possuem metodologia capaz de validar associação causal. Além disso, a metodologia de diagnóstico não é padronizada uma vez que são a descrição do caso e não envolvem experimentos.

Importante destacar que geralmente são publicados aqueles casos que tiveram sucesso com a intervenção, o que constitui um viés de publicação. Editores desprestigiam relatos e séries de casos, somente publicando aqueles realmente relevantes e que acrescentem avanços a determinado assunto.

Redação do relato de caso e da série de casos

O periódico intitulado *Cases* (casesjournal.biomedcentral.com) é inteiramente dedicado a relatos de casos clínicos únicos. Também existem guias para redação de relato de caso e série de casos, como o *Consensus-based Clinical Case Reporting Guideline Development (The CARE Guidelines)* (www.equator-network.org/reporting-guidelines/care).

Análise crítica do relato de caso e da série de casos

Podemos destacar algumas questões relevantes para a análise crítica de estudos de caso e série de casos:

1. o consentimento informado foi obtido? Há aprovação de Comitê de Ética em Pesquisa? Há a descrição de riscos associados com a intervenção? Devemos tomar todos os cuidados para apresentar os casos da forma ética e contributiva para o manejo de determinada condição de saúde, limitando-se a concluir somente o que for possível ao desenho de estudo. Os relatos de caso são eticamente aceitáveis desde que respeitados os preceitos relacionados à privacidade dos participantes, à confidencialidade dos dados e a dignidade humana. O relato de caso não é isento de riscos, podendo ocorrer quebra da confidencialidade. Esta pode trazer danos, materiais e morais, ao participante e a terceiros. Não é permitida qualquer forma de identificação do participante sem o seu consentimento. Qualquer informação que possibilite a identificação deve ser evitada, tais como: nome, codinome, iniciais, registros individuais, informações postais, números de telefone, endereços eletrônicos, fotografias, figuras, características morfológicas, entre outros. Sempre que o relato de caso requerer o uso de imagem do participante, deverá ser obtida a autorização do uso de imagem no Termo de Consentimento Livre Esclarecido (TCLE) ou em documento separado, preservando-se a autoria de quem coletou a imagem, nos termos da lei. O consentimento do participante (ou responsável legal) para a elaboração do relato de caso é essencial e deve ser obtida previamente à publicação ou divulgação, por meio de

TCLE, acompanhado do termo de assentimento quando necessário. A dispensa do TCLE será analisada pelo Comitê de Ética em Pesquisa, mediante apresentação de justificativa pertinente. Conflitos de interesse devem ser declarados no manuscrito;

2. os critérios diagnósticos são claramente definidos? Os critérios de inclusão e exclusão são citados? Os critérios diagnósticos utilizados devem ser acurados (sensíveis e específicos). Queixas atuais, gravidade e limitação pela doença, anormalidades do exame físico, alterações dos parâmetros laboratoriais e diagnósticos devem ser relatados. Deve-se excluir causas secundárias da condição de saúde ou doença relatada, a fim de evitar potenciais confundidores. Características prognósticas, doenças paralelas e tratamentos adicionais também devem ser mencionados;

3. os detalhes da intervenção são descritos? A intervenção deve ser descrita fidedignamente, com a dosagem, mudanças da dosagem, frequência de aplicação, duração, momento de início, fim, mudança, pausa, bem como acontecimento esperado, tolerância e efeitos colaterais. Os motivos nas quais se baseiam o tratamento e outras medidas relevantes também devem ser relatadas. Tratamentos paralelos também devem ser mencionados. No caso de tratamento cirúrgico, este também deve ser descrito minuciosamente para poder ser reproduzida por outros pesquisadores.

4. os desfechos clínicos são relevantes e claramente definidos? O desfecho deve objetivar a evolução (prognóstico). Descrever a evolução da doença e correlação temporal exata e transparente com o tratamento executado (se possível com documentação de imagem, fluxograma e tabelas com a evolução do quadro). Devem ser definidos os parâmetros importantes de evolução e pontos finais específicos;

5. há a descrição da percepção do paciente quanto ao desfecho e à intervenção efetuada? Se possível, deve-se relatar a evolução de queixas, limitação de funções e percepção do paciente quanto ao desfecho e intervenção realizada;

6. a conclusão é compatível com as limitações de um relato de caso? O relato de caso deve resumir os principais aspectos do seu caso, justificar a sua singularidade ou raridade, sugerir recomendações e apontar as conclusões. Nas conclusões, o autor deve ser cuidadoso, uma vez que se trata de um ou poucos casos relatados, sem poder estatístico para estabelecer uma evidência científica forte.

Estudos pré-clínicos

Talvez a primeira questão seja: qual a relevância dos estudos pré-clínicos? Os estudos pré-clínicos são amplamente utilizados para geração de conhecimento e compreensão de fenômenos naturais e sociais, além de terem surgido para resolver as curiosidades intrínsecas do ser humano. Além disso, fornecem subsídios para a pesquisa aplicada, provendo os alicerces teóricos/experimentais que possibilitam a proposição de estudos que visem avanços práticos para a sociedade.

Os métodos alternativos não clínicos, como animais de experimentação, modelos biológicos (cultura celular, tecidos), modelos físicos e/ou mecânicos (pulmão, coração, softwares, simuladores), podem responder muitas perguntas e indicar caminhos durante o desenvolvimento de pesquisas científicas. Contudo, deve ser considerada a complexidade inerente de cada modelo. Existem quatro grandes categorias de experimentos: estudos *in silico*, estudos *in vitro*, estudos *ex vivo* e estudos *in vivo*. Cada tipo desses estudos tem particularidades e compreender as particularidades dos tipos de estudo oferece uma visão sobre a validade das conclusões que se pode obter.

Independentemente do tipo de estudo, este apresenta uma pergunta de pesquisa definida e tem um objetivo principal, que orienta a escolha da espécie ou do modelo animal, da linhagem celular ou do modelo *in silico* a ser utilizado. É imprescindível formular uma justificativa que seja plausível, fundamentada em bases teóricas e que tenha respaldo ético, quando aplicável, no momento de definir o modelo experimental a ser utilizado. Além disso, é possível trabalhar com esses ensaios alternativos de forma associada, pois as informações que elas fornecem podem ser complementares.

A escolha de qualquer modelo experimental deve seguir o objetivo de pesquisa, e o método desenvolvido deve ser claro, detalhado e reprodutível. Um dos principais problemas que resultam de uma metodologia descrita de forma não adequada e com poucos detalhes, são as possíveis falhas e resultados incongruentes dentro das condições de outro laboratório quando se tenta reproduzir o modelo. Então, a descrição dos grupos em estudo, das dosagens e concentrações de medicamentos, do tamanho amostral, dos experimentos, dos reagentes (e fabricantes) e as demais informações pertinentes deve ser clara. É importante, também, disponibilizar as referências dos estudos que desenvolveram aquele método, quando elas estão publicadas.

Estudos *in silico*

Os estudos *in silico* referem-se a experimentos biológicos realizados inteiramente via simulação em computador. Por exemplo, estudos que preveem como medicamentos interagem com o corpo e com os patógenos. A expressão foi usada pela primeira vez em 1989, pelo matemático Pedro Miramontes. Através dos ensaios *in silico* pode-se prever possíveis efeitos e interações de moléculas (sintéticas ou biológicas). Apesar de serem uma nova alternativa aos ensaios *in vivo*, os ensaios *in silico* limitam-se pelo fato de realizarem uma previsão do que pode ocorrer.

Os ensaios *in silico* consistem, resumidamente, em três modelos, que são os moleculares, os de sequenciamento bacteriano e os simuladores de células vivas. Os modelos moleculares demonstram como as drogas e outras substâncias interagem com os receptores nucleares das células; os modelos de sequenciamento bacteriano são uma alternativa aos métodos *in vitro* de identificação de bactérias, que sequenciam o DNA e o RNA bacterianos e; os modelos simuladores de células vivas simulam com precisão o comportamento das células vivas.

Essas ferramentas bioinformáticas têm um importante papel na aceleração do desenvolvimento de fármacos. O uso da tecnologia para modelos de previsão computacionais tem crescido e apresenta uma ampla possibilidade de testagem pré-clínica.

Estudos *in vitro*

Do latim, “dentro do vidro”, os ensaios *in vitro* representam os ensaios biológicos realizados fora de sistemas vivos, no ambiente controlado e fechado de um laboratório experimental e que são feitos normalmente em recipientes, como placas. Ensaios *in vitro* aplicam-se apenas à linhagem celular utilizada. Por exemplo, ensaios realizados em culturas de células cancerígenas para identificar substâncias anticancerígenas.

Os ensaios *in vitro* são utilizados como alternativa aos ensaios *in vivo*, pois permitem uma análise mais específica, simples, conveniente e detalhada do que pode ser visto no ensaio *in vivo*. Como replicam as condições celulares precisas de um organismo, os estudos *in vitro* podem levar a resultados que não correspondem às circunstâncias que ocorrem em torno de um organismo vivo, podendo não prever de maneira completa ou precisa os efeitos. Os pesquisadores devem ter cuidado para evitar interpretações exageradas.

O cultivo celular compreende um conjunto de técnicas que permitem manter células e tecidos *in vitro*, ou seja, em um microambiente controlado, conservando ao máximo suas propriedades fisiológicas, bioquímicas e genéticas. A cultura primária é aquela feita diretamente com tecidos de um organismo (utilizando-se desagregação mecânica ou enzimática para obter as células), que permite manter todas as características e funções do tecido de origem, obtendo-se células heterogêneas. A cultura secundária é feita a partir de células obtidas de uma cultura primária. Estas células irão proliferar por um número finito de divisões celulares e depois sofrem senescência, morrendo por apoptose.

Uma linhagem celular contínua é aquela em que as células derivam de culturas primárias ou de linhagens celulares que podem se dividir indefinidamente, devido a um processo conhecido como imortalização. A imortalização pode ser espontânea ou induzida, possui menor variação dos resultados e possibilita uma maior reprodutibilidade. É possível obter também células transformadas, que são modificadas genética e morfológicamente. Tais células podem sofrer a transformação em cultura, utilizando-se substâncias químicas, agentes virais ou físicos como a luz ultravioleta.

Estudos *ex vivo*

Do latim, “fora do vivo”, os ensaios *ex vivo* são realizados no ou sobre o tecido de um organismo em um ambiente externo com mínima alteração das condições naturais, permitindo a experimentação em células ou tecidos de um organismo sob condições mais controladas do que é possível em experimentos *in vivo*. Os ensaios *ex vivo* aplicam-se apenas ao organismo que fornece a célula. Por exemplo, estudo da permeabilidade dos fármacos utilizando-se epitélio intestinal de rã ou ensaio de membrana corioalantóica (CAM) em pintinhos, em que a angiogênese é promovida na CAM de um embrião de galinha fora do organismo (galinha).

Através dos ensaios *ex vivo*, consegue-se realizar testes ou ensaios que de outra forma não seriam possíveis ou éticos em indivíduos vivos, como medições de propriedades físicas, térmicas, elétricas, mecânicas, ópticas e outras propriedades do tecido, especialmente em vários ambientes que podem não ser sustentáveis, como por exemplo, em pressões ou temperaturas extremas. São amplamente utilizados para o desenvolvimento de procedimentos cirúrgicos. Limita-se em não replicar as condições do organismo como um todo.

Estudos *in vivo*

Do latim, “no vivo”, os estudos *in vivo* correspondem aos ensaios realizados em animais vivos no ambiente controlado e geralmente fechado de um laboratório de experimentação, como por exemplo, um ensaio de edema de pata em ratos para identificar substâncias antiedemato-gênicas. Os estudos *in vivo* são frequentemente empregados ao invés dos estudos *in vitro* por serem mais adequados para observar os efeitos gerais de um experimento em um organismo vivo. Apesar disso, podem levar a resultados que não correspondem às circunstâncias reais do ser humano.

Existem diversos modelos animais utilizados em pesquisa básica e atividades de ensino. Os animais de experimentação são criados e mantidos em ambientes monitorados, visando garantir a reprodutibilidade dos resultados, através do controle sanitário, ambiental e do *status* genético. Alguns exemplos são o Zebrafish (*Danio rerio*), empregado em estudos genéticos, toxicológicos e

relacionados a processos regenerativos; anfíbios (*Xenopus laevis*), utilizados em estudos relacionados com o sistema oscilatório endógeno e funcionamento do sistema muscular; cães (*Canis familiaris*), em estudos comportamentais e biomédicos associados com doenças genéticas e neuromusculares; coelhos (*Oryctolagus cuniculus*), usados em estudos imunológicos relacionados à produção de vacinas; etc.

Mas, os roedores correspondem ao modelo mais utilizado na experimentação animal, nomeadamente em estudos que visam caracterizar padrões neurais relacionados com diferentes comportamentos, estudos que avaliam interações sistêmicas durante respostas de defesa, e estudos farmacológicos relacionados à eficácia de determinados tratamentos. Isso se deve a algumas características que os tornam vantajosos em pesquisas bem delineadas. Dentre estas características, destaca-se a facilidade de criação e manejo; anatomia, fisiologia e genética bem conhecidas e com certo grau de similaridade com o ser humano; existência de linhagens com alterações genéticas e fisiológicas espontâneas; além da disponibilidade de linhagens geneticamente modificadas.

As linhagens com alterações gênicas e fisiológicas espontâneas são aquelas em que as mutações ocorrem ao acaso e conferem aos animais a propensão para o desenvolvimento de determinadas doenças, como por exemplo, ratos SHR (*Spontaneously hypertensive rat*), utilizados em estudos relacionados ao funcionamento do sistema cardiovascular; camundongos *Nude* (atímicos), ideais para estudos relacionados à imunidade inata, muito utilizados em estudos com tumores heterotransplantados e; camundongos espontaneamente obesos, empregados em estudos relacionados ao metabolismo energético.

Os modelos geneticamente modificados, cujos genomas foram alterados através da engenharia genética, adquirem fenótipos de interesse, como os animais transgênicos, que expressam genes não presentes em seu genoma ou superexpressam genes de interesse, e os animais *Knockout/Knockin*, que tiveram a deleção ou inativação de um gene de interesse, por um processo sistêmico ou restrito às células ou a tecidos-alvos.

O uso de animais em laboratório põe em cheque princípios bioéticos, em que a utilização de qualquer modelo requer uma justificativa do porquê de ele ser relevante para a proposta da pesquisa. Ou seja, a utilização consciente de animais em experimentação científica e de ensino. A lei Arouca (nº 11.794, de 08 de outubro de 2008) criou o Conselho Nacional de Controle de Experimentação Animal (CONCEA), estabeleceu procedimentos para uso de animais em atividades de ensino e pesquisa, organizou o Sistema Nacional de Experimentação Animal e trouxe orientações às Comissões de Ética no Uso de Animais (CEUAs).

O CONCEA é a autoridade legal brasileira responsável por regular o uso de animais. Ele é responsável pelo credenciamento de instituições para criação e utilização de animais em pesquisa, atua na formulação e revisão das normas e dos procedimentos para instalação e manutenção dos centros de experimentação animal, monitora o cumprimento das regras e diretrizes para o uso de animais de forma ética e humanitária no ensino e na pesquisa, entre outras atribuições.

A utilização de animais em atividades educacionais é restrita a estabelecimentos de ensino superior e estabelecimentos de educação profissional técnica de nível médio da área biomédica. São consideradas como atividades de pesquisa científica todas aquelas relacionadas com ciência básica, ciência aplicada, desenvolvimento tecnológico, produção e controle da qualidade de drogas, medicamentos, alimentos, imunobiológicos, instrumentos, ou quaisquer outros testados em animais, conforme definido em regulamento próprio. Não se incluem nessa categoria as atividades práticas zootécnicas relacionadas à agropecuária, como por exemplo, a profilaxia e tratamento veterinário de animais que deles necessitem; o anilhamento, a tatuagem, a marcação ou a aplicação de outro método com finalidade de identificação do animal, desde que cause apenas dor ou aflição momentânea ou dano passageiro; e as intervenções não-experimentais relacionadas às práticas agropecuárias.

A norma se aplica aos animais das espécies classificadas como Filo Chordata, cujos animais apresentam notocorda, fendas branquiais na faringe e tubo nervoso dorsal único na fase embrionária, e subfilo Vertebrata, cujos animais cordados que

têm, como característica exclusiva, um encéfalo grande encerrado numa caixa craniana e uma coluna vertebral.

Redação de estudos pré-clínicos

Existem guias que orientam a construção e elaboração da metodologia de pesquisas. Por exemplo, a diretriz ARRIVE (*Animal Research: Reporting of In Vivo Experiments*), disponível na rede EQUATOR (www.equator-network.org/reporting-guidelines/improving-bioscience-research-reporting-the-arrive-guidelines-for-reporting-animal-research), contém o *The ARRIVE Essential 10*, que é uma lista com dez itens de recomendação para elaboração de estudos com animais, e o *The ARRIVE Guidelines 2.0*, que apresenta dez itens adicionais. Outro guia, o PREPARE (*Planning Research and Experimental Procedures on Animals: Recommendations for Excellence*) (pubmed.ncbi.nlm.nih.gov/28771074), é um *checklist* sobre como planejar experimentos com animais com o objetivo de garantir a qualidade dos mesmos.

O MIQE (*Minimum Information for Publication of Quantitative Real-Time PCR Experiments*) (www.gene-quantification.de/miqe-press.html) é um guia voltado para experimentos que utilizam a técnica de PCR em tempo real e que procura aumentar a transparência dos desenhos experimentais e garantir a reprodutibilidade e consistência dos resultados entre laboratórios. De forma similar, o MIAME (*Minimum Information About a Microarray Experiment*) (www.nature.com/articles/ng1201-365) descreve como registrar e relatar ensaios com expressão gênica por *microarrays* de forma padrão, para que o acesso aos dados e a interpretação dos resultados seja facilitado para a comunidade. Existem vários guias similares disponíveis e que são úteis para a redação e o desenvolvimento dos ensaios.

Análise crítica dos estudos pré-clínicos

Devido as suas particularidades e especificidades, é difícil estabelecer perguntas que norteiem a análise crítica de estudos pré-clínicos, pois sua qualidade depende do tipo de modelo experimental utilizado e dos experimentos. Porém, podemos avaliar um estudo pré-clínico com base em questões comuns:

1. a pergunta de pesquisa é original, ética e interessante? A pergunta de pesquisa deve ser inédita e relevante para contribuir com o conhecimento científico, além de ser plausível e ética. Deve estar apoiada a uma base teórica coerente e lógica, para servir de justificativa para o desenvolvimento do estudo;
2. o modelo experimental utilizado foi o mais adequado? A descrição da metodologia deve mostrar que o modelo escolhido, efetivamente, gera resultados que comprovam as hipóteses de estudo. A escolha do modelo deve ser assertiva e justificada;
3. o método foi claramente descrito? O método deve ser descrito de forma detalhada, clara e cronológica. Deve-se reportar todas as informações pertinentes do estudo como espécie, linhagem, grupos (controle e experimentais), características basais (massa corporal, sexo, idade, quantidade de células), meios de cultura, dieta, doses, soluções, reagentes (pureza, fabricante), tempo de tratamento (agudo/crônico), e qualquer outra informação importante que permita a reprodutibilidade do experimento;
4. os princípios éticos, quando aplicáveis, foram respeitados? O cuidado e o manejo ético dos animais foi realizado? Métodos alternativos foram considerados? Basicamente, isso envolve o princípio dos 3 Rs de Russel e Burch (*Reduce, Replace and Refine*), que propõem a redução do número de pesquisas realizadas em modelos animais, redução do número de animais utilizados e aumento na qualidade do tratamento estatístico, a substituição do uso de animais por métodos alternativos e o refinamento das técnicas utilizadas, visando diminuir a dor e o sofrimento dos animais, incluindo cuidados de analgesia e assepsia nos procedimentos. A reflexão sobre a necessidade da utilização desses modelos é fundamental, considerando o avanço da tecnologia e os diversos modelos alternativos existentes;
5. os resultados foram adequadamente analisados e descritos? Modelos *in vivo* devem ter amostra suficiente para todos os grupos, considerando possíveis perdas, que devem ser reportadas quando ocorrem (investigação da perda do espécime). Diferenças basais, como massa corporal entre os grupos, devem ser consideradas nos modelos na

análise estatística, quando pertinente. Em ensaios *in vitro*, contaminação de cultura celular e inviabilidade celular devem ser reportadas, e os experimentos devem ser realizados em triplicata. Esses são aspectos que orientam a qualidade metodológica dos estudos e contribuem para a apresentação de resultados fidedignos.

6. e, finalmente, as conclusões apresentadas corroboram para responder a pergunta de pesquisa? A conclusão tem que fazer referência aos objetivos da pesquisa, independentemente da metodologia utilizada.

Referências

Bonita, R.; Beaglehole, R.; Kjellström, T. Epidemiologia básica. 2. ed. São Paulo: Santos, 2010-2019. 213 p.

Greenhalgh, T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Kinele, G.S. & Kiene, H. Como escrever um relato de caso. Tradução de Bernardo Kaliks com autorização dos autores e da Revista Der Merkurstab, do original em alemão: Methodik der Einzelfallbeschreibung. Der Merkurstab, 62(3): 239-42, 2009.

Lei Nº 11.794, de 8 de outubro de 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/l11794.htm.

Ministério da Saúde. Carta Circular 166/2018-CONEP/SECNS/MS.

Parente, R.C.M.; Pinho de Oliveira, M.A.; Celeste. R.K. Relatos e Série de Casos na Era da Medicina Baseada em Evidência. Bras. J. Video-Sur. 2010;3(2):67-70.

Pinho de Oliveria, M.A.; Velarde, G.C.; Moreira de Sá, R.A. Entendendo a pesquisa clínica V: relatos e séries de casos. FEMINA. 2015;43(5):235-238.

Russel, W.M.S. & Burch, R.L. The Principles of Humane Experimental Technique. London: Methuen. 1959.

Pesquisa qualitativa

*Cristiana Palma Kuhl
Luciana Stein e Silva
Edison Capp
Otto Henrique Nienov*

A pesquisa qualitativa, também chamada de pesquisa interpretativa, foi durante muitos anos um tipo de investigação exclusivo da área das ciências humanas. É um tipo de pesquisa que não expressa números, ou seja, não é quantificada e, infelizmente, evidências que contêm números são mais facilmente aceitas como fato quando quantificáveis do que se não o forem. Atualmente, a pesquisa qualitativa é cada vez mais reconhecida como complementar.

Apesar da pesquisa qualitativa não se enquadrar na pirâmide de hierarquia das evidências, quando bem delineada e bem conduzida, é tão válida e necessária quanto a pesquisa quantitativa, fornecendo um achado de grande qualidade e, inclusive, servindo de suporte e apoio para o desenvolvimento da pesquisa quantitativa por mostrar lacunas no conhecimento que necessitam ser preenchidas.

As possibilidades da metodologia qualitativa em projetos de pesquisa na área da saúde não implicam a desvalorização da metodologia quantitativa, como se um tipo de metodologia se construísse em oposição ao outro. Muito pelo contrário. São metodologias complementares à medida que respondem perguntas de pesquisa diferentes: enquanto que a pesquisa quantitativa se utiliza do raciocínio dedutivo, a pesquisa qualitativa recorre ao raciocínio indutivo. Além disso, muitas vezes, a pesquisa qualitativa é um pré-requisito necessário para se conduzir uma pesquisa quantitativa.

O que é pesquisa qualitativa?

A pesquisa qualitativa, por natureza, não é padronizada, não é confinada e depende da experiência subjetiva tanto do pesquisador como do sujeito pesquisado. É interpretativa e realizada em profundidade e não é um procedimento técnico enumerativo. A pesquisa qualitativa explora o que necessita ser explorado e molda-se de acordo.

Os métodos qualitativos de pesquisa são muito úteis para pesquisar um território desconhecido, não tendo qualquer utilidade na mensuração de fenômenos em grandes grupos, sendo basicamente úteis para quem busca entender o contexto de algum fenômeno. Permitem a observação de vários elementos simultaneamente em um pequeno grupo de indivíduos, propiciando conhecimento aprofundado de um evento e a explicação de comportamentos daqueles indivíduos.

A pesquisa qualitativa visa examinar aspectos específicos, particulares, em uma situação natural, aplicado a um grupo específico de pessoas, com abordagem ampla, aberta e profunda, de forma a esclarecer alguns aspectos, gerar ou modificar hipóteses e priorizar áreas a investigar. Resumidamente, a pesquisa qualitativa tem interesse em saber como as pessoas veem e como se sentem quando defrontadas com as situações estudadas, tentando encontrar o sentido de, ou interpretar, fenômenos em termos dos significados que as pessoas atribuem a eles.

Definição da pergunta de pesquisa

A pesquisa qualitativa exige dos pesquisadores um período de reflexão e planejamento da investigação. A definição do tema de pesquisa é parte fundamental do projeto. O tema é o assunto que se quer investigar, sempre amplo e não traz, necessariamente, um problema científico. Entretanto, a explicitação temática a ser investigada permite situar o leitor dentro do contexto mais amplo no qual se insere o objeto que será investigado. Neste sentido, a definição do tema fornece importantes subsídios para a justificativa da própria pesquisa.

Por muitas vezes, não se tem uma pergunta de pesquisa estabelecida no início da investigação qualitativa, ou seja, não se tem uma hipótese formulada. A pergunta de pesquisa é modulada ao longo da coleta dos dados e a hipótese acaba sendo formulada à medida que os dados são coletados. Independente disto, é essencial planejar a pesquisa qualitativa e refletir em como abordar sua questão de pesquisa, podendo ser montada no acrônimo PICo (Figura 1).

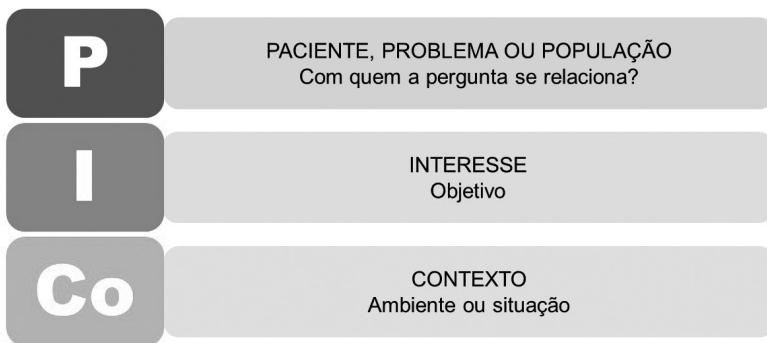


Figura 1. Acrônimo PICo.

Formulação da hipótese

A partir da definição do tema, realiza-se a coleta dos dados, através de observações, entrevistas, documentos, mensagens eletrônicas, etc. e, a medida que são coletados os dados, constrói-se a hipótese de pesquisa, utilizando o que é conhecido como raciocínio indutivo.

Desta forma, a hipótese pode surgir quando os procedimentos investigados estão em plena evolução, ou seja, a hipótese definitiva pode não ter sido elaborada até que o estudo esteja em andamento. Porém, é nessas circunstâncias que o pesquisador deve se assegurar de que tenha, desde o início, delineado cuidadosamente um foco particular de pesquisa e identificado algumas questões específicas para tentar responder.

Objeto de estudo

O objeto de estudo corresponde ao problema que se quer investigar e é fundamental que seja bem delimitado. Há a possibilidade de redefinir o objeto de estudo durante o próprio período de desenvolvimento da pesquisa. Essa redefinição pode se dar em função das especificidades do grupo e/ou local de estudo, das mudanças sociais, econômicas, políticas ou em razão mesmo da pertinência do objeto para o grupo de estudo.

A construção do objeto de estudo ocorre da seguinte maneira:

1. parte-se de um tema de pesquisa amplo;
2. busca-se na literatura, tanto científica como outras fontes de informação (internet, jornais, etc.), sobre o tema;
3. identifica-se, com base na revisão da literatura, as especificidades do tema, os aspectos privilegiados pelos diferentes enfoques, a atualidade da questão e sua relação com a realidade empírica;
4. coloca-se, a partir do material consultado e de observações realizadas, um conjunto de perguntas para o tema que se quer investigar;
5. confronta-se essas perguntas com as questões colocadas pela área de conhecimento específica da investigação;
6. escolhe-se um aspecto do tema que seja de interesse do pesquisador e que tenha relação com uma determinada realidade empírica e com as questões colocadas pela área de conhecimento e;
7. elabora-se a problemática de investigação, delimitando o tema, as questões a serem investigadas e sua relação com a área de conhecimento.

Elaboração dos objetivos

Em alguns modelos de projeto de pesquisa, a definição das questões a serem investigadas aparecem no formato de objetivo principal e objetivos específicos.

Existe amostra na pesquisa qualitativa?

Na pesquisa qualitativa não se trabalha com a ideia de “amostra”, visto que isso traz implícita a concepção de que uma determinada parte da população é representativa do todo. O que se quer é avaliar um grupo de sujeitos para uma determinada pergunta que não é quantificável. Portanto, as pesquisas qualitativas são planejadas para fornecer uma visão de dentro do grupo de pesquisa, uma visão conhecida como *êmica*.

Trabalha-se com um elevado número de questões e, para que isso seja possível, busca-se estudar sempre um grupo pequeno de pessoas, o qual é escolhido de acordo com critérios previamente definidos conforme os objetivos de estudo. Para que seja feita a escolha, é necessário o conhecimento prévio do contexto a ser investigado, o que é feito com a ajuda de uma série de técnicas de pesquisa dentro de uma orientação etnográfica.

A coleta de dados depende do estabelecimento de uma relação entre o pesquisador e o sujeito pesquisado, sendo frequente o uso de uma técnica de obtenção de participantes conhecida como técnica de *snowball*, pela qual cada novo participante indica novos participantes para fazerem parte da pesquisa. A técnica de *snowball* garante a seleção de indivíduos do mesmo grupo que se está querendo investigar.

Etnografia

A etnografia (Figura 2) refere-se a escrita (estudo) sobre determinadas pessoas em locais ou “no campo” que ocorrem naturalmente, através de métodos de coleta de dados que captam seus significados sociais e suas atividades comuns, envolvendo a participação direta do pesquisador no local. O principal foco da etnografia é o estudo da cultura e o comportamento de determinados grupos sociais.

Uma abordagem qualitativa aos problemas de saúde identifica-se de várias formas com o método etnográfico. O método etnográfico é diferente de outros modos de fazer pesquisa qualitativa, pois segue alguns princípios: pesquisa de campo (conduzida no local em que as pessoas convivem e socializam),

multifatorial (conduzido pelo uso de duas ou mais técnicas de coleta de dados), indutiva (acúmulo descritivo de detalhe) e holística (retrato mais completo possível do grupo em estudo).

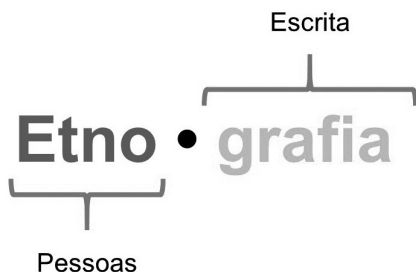


Figura 2. Definição de etnografia.

Estudo *in loco*

Na pesquisa qualitativa é preciso entender o ponto de vista do “nativo”, procurando estabelecer e conhecer os seus significados. O trabalho do pesquisador deve ser o de examinar minuciosamente os diversos aspectos da vida dos diferentes grupos sociais. Por isso, o trabalho de campo intensivo com observações *in loco* é fundamental, pois nos conduz ao ponto de vista do sujeito em seu ambiente natural.

A pesquisa de campo do tipo etnográfica implica uma relação social entre o pesquisador e os sujeitos pesquisados. A qualidade dos dados obtidos depende da maneira como essa interação social se estabelece, pela relação de proximidade, de confiança, baseada em princípios éticos. A aceitação do pesquisador por parte dos sujeitos é fundamental para a qualidade do dado obtido. Inclusive, a presença do pesquisador pode influenciar na observação.

Neste aspecto, o pesquisador enfrenta o problema de transitar entre dois mundos: o acadêmico-científico e o do grupo pesquisado. O pesquisador precisa alcançar o equilíbrio, de modo que tenha uma boa inserção no ambiente, não sendo influenciado pelo ponto de vista do grupo que está sendo investigado. Ou seja, deve-se realizar a investigação do problema, e não fazer o relato de acordo com o ponto de vista dos indivíduos.

Universo empírico

Dentro da pesquisa qualitativa, o universo empírico refere-se ao grupo que será estudado e ao local onde será realizada a investigação. A escolha deve dar-se em função do objeto de investigação, ou seja, o local e o grupo de indivíduos deve apresentar as melhores condições de explicitação do problema de pesquisa. É fundamental que o pesquisador busque, dentro de seu universo, a maior diversidade possível, de forma a obter uma saturação dos dados adequada.

Os aspectos que devem ser considerados para se determinar o universo empírico são:

1. grupo(s) de estudo: pode ser delimitado em função de diferentes critérios, como faixa etária, classe social, condição física, sexo, gênero, profissão, condição marital, etc.;
2. local: a localidade ou região em que os indivíduos serão recrutados deve garantir as melhores condições de acesso ao grupo, tanto no que se refere à disponibilidade em grande número do perfil desejado, como no que diz respeito à aceitação da presença do pesquisador por parte dos membros do grupo;
3. temporalidade: é fundamental explicitar qual o período de tempo que será contemplado no estudo. O recorte temporal deve estar, evidentemente, em conformidade com o objeto de investigação e;
4. número de pessoas: não é determinado por cálculo amostral, tampouco rígido. É indicado pela saturação ou recorrência de dados, isto é, o momento em que a busca por novos participantes não acrescenta mais nenhum dado novo ou inédito à investigação.

Métodos de coleta de dados

Uma vez estabelecido o objeto de pesquisa, o passo seguinte é decidir como coletar as informações para estudá-lo. No método, simplificarmente, a observação consiste em ouvir o que as pessoas têm a dizer e explorar as ideias e preocupações que os próprios sujeitos trazem.

Os métodos de pesquisa qualitativa permitem e, na verdade, exigem a modificação da questão de pesquisa à medida que os achados são gerados, uma técnica conhecida como foco progressivo. Para isso, utiliza-se uma abordagem interativa, que consiste em alterar os métodos de pesquisa e a hipótese à medida que a coleta de dados progride. A utilização da abordagem iterativa não descredita os resultados da pesquisa qualitativa. Na verdade, mostra sensibilidade elogiável à riqueza e à variabilidade do assunto.

Diversas são as técnicas de pesquisa qualitativa (Quadro 1) que podem ser utilizadas para coleta de dados, e estas devem ser pensadas e escolhidas conforme os objetivos do estudo. O uso de técnicas combinadas, ou seja, a utilização coerente de mais de uma técnica permite suprir as lacunas e tornar mais completa a coleta de informações.

Quadro 1. Técnicas de coletas de dados da pesquisa qualitativa. Adaptado de Greenhalgh T. (2015) e Victoria CG, Knauth DR e Hassen MNA (2000).

Técnica de coleta de dados	Descrição
Etnografia (observação participativa ou direta)	Observação na qual o pesquisador também ocupa uma função ou faz parte da situação, além de observar (dualidade da observação). Examina-se um evento, um grupo de pessoas, um indivíduo dentro de um contexto, com o objetivo de descrevê-lo. Deve-se levar em consideração o efeito da presença do observador na observação.
Entrevista	<p>É importante ter clareza nos objetivos da pesquisa. São os objetivos que irão definir quem entrevistar, o conteúdo das entrevistas, o número de pessoas entrevistadas, o número de entrevistas com cada sujeito e o tipo de entrevista apropriada:</p> <p>Semiestruturada: conversa frente a frente, ou por telefone, com o objetivo de explorar assuntos ou tópicos em detalhes. Utiliza-se uma ampla lista de questões ou tópicos (questionário padronizado), conhecida como guia por tópicos.</p> <p>Narrativa (sem roteiro): Entrevista realizada de modo menos estruturado (informal), com o objetivo de obter uma longa história junto ao entrevistado (em geral, uma história de vida ou a história de como a doença se desenrolou ao longo do tempo). O pesquisador entrevistador contém-se para não dizer mais do que "fale-me mais sobre isto".</p>

Quadro 1. Técnicas de coletas de dados da pesquisa qualitativa. Adaptado de Greenhalgh T. (2015) e Victoria CG, Knauth DR e Hassen MNA (2000) (continuação).

Técnica de coleta de dados	Descrição
Grupos focais	É fundamental que haja um foco, um tópico a ser explorado. O método de entrevista em grupo inclui e utiliza a interação grupal para gerar dados. Um grupo focal pode abordar um tema específico (captar diferentes visões sobre o mesmo), um grupo (captar a visão de mundo ou de determinados temas) ou ambos (entender a profundidade de um comportamento dentro de um grupo determinado). É realizada em três fases: I) convite aos participantes e preparação do encontro (roteiro); II) encontro e mapeamento dos participantes e; III) transcrição e compilação dos dados.
Análise de discurso ou de documentos	Pouco utilizada na pesquisa qualitativa. Corresponde ao estudo detalhado das palavras, sentenças e formatos utilizados em contextos sociais determinados. Inclui o estudo de conversas decorrendo naturalmente, além de material escrito, como documentos oficiais (leis, regulamentos), pessoais (cartas, diários, autobiografias) e públicos (livros, jornais, revistas, discursos).
História da vida	Compreender o desenvolvimento da vida do sujeito investigado e traçar uma biografia que descreva sua trajetória até o momento atual. A história de vida pode, além de recuperar as experiências dos indivíduos, recolher também crenças, mitos, tradições, o que permite o melhor entendimento da própria história e trajetória dos participantes. A coleta de história de vida é normalmente realizada ao longo de vários encontros, quando o pesquisador e o sujeito pesquisado desenvolvem uma relação de confiança. Também deve-se considerar a sequência dos eventos.
Rede de relações	Conjunto de vínculos entre um conjunto específico de pessoas, e as características desse conjunto. Ao pesquisador interessa a organização dessas redes, os intercâmbios realizados (formais ou não, explícitos ou implícitos) e as formas de troca socialmente aceitáveis. Importa observar o que é trocado (conteúdo de vínculos), com quem é trocado (relações horizontais ou verticais) e quanto é trocado (densidade dos vínculos).
Elaboração de desenhos	Propor aos sujeitos pesquisados que representem graficamente uma determinada situação ou concepção. Esta técnica é apropriada em casos em que a comunicação oral não se mostra suficiente para levantar as impressões do sujeito pesquisado. A partir do desenho, o pesquisador inicia uma discussão com o sujeito pesquisado, apoiada nos elementos que se encontram no desenho.
Classificação/ ordenação de fotos/gravuras	Consiste na apresentação, aos sujeitos pesquisados, de um número determinado de fotos de pessoas em diferentes situações ou com diferentes condições físicas, solicitando que os sujeitos pesquisados as coloquem em ordem quanto ao estado de saúde (do mais saudável ao mais doente, por exemplo).

Após a coleta de dados, estes podem ser registrados de duas formas (Quadro 2).

Quadro 2. Técnicas de registro de dados de pesquisa qualitativa. Adaptado de Victoria CG, Knauth DR e Hassen MNA (2000).

Técnica de registro	Descrição
Diário de campo "eletrônico"	É um documento pessoal do pesquisador e, nada mais é do que um registro fiel e detalhado de cada visita a campo, independentemente das técnicas de pesquisa utilizadas. Muitas vezes, fornecem subsídios para analisar os dados coletados.
Síntese de dados	Corresponde a uma primeira escrita na qual se busca uma sistematização inicial dos dados e não deve ser confundida com o relatório final da pesquisa. É uma parte intermediária entre a coleta e a análise de dados. Suscintamente, corresponde ao início da "lapidação" dos dados.

Com isso, obtêm-se diferentes tipos de dados: registros escritos, observacionais e discursivos, os quais são complementares e podem ser utilizados na triangulação de informações, permitindo uma aproximação mais ampla da realidade. Ao final, o pesquisador terá disponível um texto volumoso a partir da transcrição de entrevistas, por exemplo. Cria-se, então, uma maneira sistemática de análise dos dados e seleciona-se passagens contendo exemplos que dão suporte à ideia que está sendo defendida no trabalho.

Validade da pesquisa qualitativa

O poder da pesquisa qualitativa está na validade (proximidade da verdade) dos dados obtidos. Isso quer dizer que, uma boa pesquisa qualitativa, usando uma seleção de métodos de coleta de dados, realmente deve atingir a essência da questão e não ser apenas superficial.

A validade dos métodos qualitativos (Figura 3) aumenta consideravelmente pelo uso de mais de um método em combinação, processo conhecido como triangulação; pelo pesquisador pensando cuidadosamente sobre o que está acontecendo e como sua própria perspectiva pode estar influenciando os dados, abordagem conhecida como reflexividade; e, por mais de um pesquisador

analisando os mesmos dados de modo independente para demonstrar a confiabilidade interexaminadores.



Figura 3. Validade dos métodos qualitativos.

No entanto, a confiabilidade interexaminadores tornou-se menos confiável como medida de qualidade em pesquisa qualitativa, devido a dois aspectos: primeiro, na maioria das pesquisas qualitativas, um pesquisador conhece os dados muito melhor que o outro, de modo que a ideia de que duas cabeças pensam melhor do que uma simplesmente não é verdadeira. Um pesquisador que foi trazido somente para conferir "temas" pode basear-se muito mais em pré-julgamentos e conjecturas pessoais do que o principal pesquisador da área. E, em segundo, não somente a "concordância" entre dois pesquisadores pode não corresponder à qualidade como também a formação similar propicia a ocorrência de vieses, de modo que alta confiabilidade interexaminador pode ser viciada.

E como saber se a pesquisa qualitativa atingiu a saturação teórica?

A saturação teórica em pesquisa qualitativa é utilizada para estabelecer ou fechar o tamanho final de uma "amostra" em estudo, interrompendo a captação de novos sujeitos. É definido como a suspensão de inclusão de novos sujeitos quando os dados obtidos passam a apresentar, na avaliação do pesquisador, uma certa redundância ou repetição de dados, não sendo considerado relevante continuar com a coleta de dados. Ou seja, não contribuiriam significativamente para o aperfeiçoamento da reflexão teórica fundamentada nos dados que estão sendo coletados.

Alguns critérios podem auxiliar o pesquisador a verificar se as informações fornecidas pelos novos sujeitos pesquisados pouco acrescentariam ao material já coletado:

1. disponibilizar os registros de dados "brutos": todos os pesquisadores têm acesso aos registros integrais de áudio e aos transcritos, desde as primeiras coletas. Dispor o suporte auditivo permite perceber elementos de entonação e prosódia e as transcrições possibilitam observações mais reflexivas e detidas sobre alguns trechos;
2. "imersão" em cada registro: são feitas leituras e audições individuais visando a identificar núcleos de sentido nas manifestações dos sujeitos participantes. Os pesquisadores "exploram" individualmente cada uma das entrevistas. Caso se esteja trabalhando com pré-categorias analíticas, os temas e enunciados correspondentes, que cada pesquisador considere importantes para instruí-las, são anotados. Se forem previstas formulações de novas categorias, cada pesquisador também as registra para depois discuti-las com os pares;
3. compilar as análises individuais (de cada pesquisador, para cada entrevista): as entrevistas são o foco, devendo-se compilar os temas e tipos de enunciados identificados em cada uma delas, por cada um dos pesquisadores do grupo (instruindo as pré-categorias ou as eventuais novas categorias). Um dos pesquisadores compila os temas e tipos de enunciados identificados por ele próprio e pelos outros investigadores. Caso sejam sugeridas novas categorias, o pesquisador compilador poderá agrupá-las, pois os colaboradores podem dar nomes diferentes a conceitos e ideias muito semelhantes. Em uma reunião, as divergências nas análises individuais são discutidas em busca de um consenso;
4. reunir os temas ou tipos de enunciados para cada pré-categoria ou nova categoria: os temas ou tipos de enunciados são o foco, em que, depois de cada entrevista ser analisada pelos pesquisadores, agregam-se as falas consideradas exemplares dos núcleos de sentido identificados. Para cada categoria são agrupados os

correspondentes temas e tipos de enunciados identificados e consensuados, fazendo constar os respectivos trechos das entrevistas que os exemplificam;

5. codificar ou nominar os dados: nomeação dos temas e dos tipos de enunciados contidos em cada pré-categoria ou nova categoria. A nomeação deve representar o conjunto de ideias, valores, sentimentos etc., abarcando as manifestações dos entrevistados que contribuem para compreender melhor o que expressaram. Nesta parte, se concentra o trabalho criativo de pesquisa, fundamentado em conhecimentos e envolvendo raciocínios indutivos: parte-se do particular (um ou vários enunciados) e chega-se a algo mais abstrato e geral (a nomeação deste conjunto);

6. alocar em uma tabela os temas e tipos de enunciados: agregando-os para cada pré-categoria ou nova categoria e destacando quando se deu a primeira ocorrência. Trata-se de uma alocação que permita a visualização dos elementos analíticos com que se trabalha;

7. constatar a saturação teórica para cada pré-categoria ou nova categoria: isso ocorre quando novos temas ou tipos enunciados não são, de maneira consistente, acrescentados após novas entrevistas. Verifica-se e soma-se a frequência do aparecimento de novos temas ou tipos de enunciados, em cada entrevista. Essa soma pode ser útil para uma constatação mais objetiva do processo de saturação. Constata-se o escasseamento de novos tipos de enunciados, correspondendo ao que se designa como saturação teórica. A coleta de novos dados por meio de novas entrevistas acrescentaria supostamente poucos elementos para discussão em relação à densidade teórica já obtida.

8. "visualizar" a saturação: pode-se transformar os dados da tabela em um gráfico, possibilitando, para cada categoria analisada, uma constatação visual da "saturação". Normalmente, utiliza-se um gráfico de barras, em que as barras correspondem aos temas e o eixo X aos sujeitos pesquisados.

Análise de dados

Considerando que são dados do tipo qualitativo, gerado a partir de observações detalhadas e entrevistas, torna-se difícil a utilização de estatística para analisar os dados. Realiza-se a interpretação subjetiva dos eventos pesquisados. Para análise de dados de imagens, existem técnicas e softwares que podem auxiliar na digitalização e avaliação de parâmetros dos desenhos.

Vantagens e desvantagens da pesquisa qualitativa

As vantagens da pesquisa qualitativa incluem a possibilidade de trabalhar com questões de valores e não se limitar às variáveis predefinidas pelo estudo, além de permitir uma compreensão sobre as opiniões dos indivíduos, pois possibilita ao pesquisador maior apropriação e aproximação dos resultados obtidos. Quando bem aplicada, a pesquisa qualitativa fornece riqueza das análises de dados e possibilita a produção de conhecimento científico, derivado de uma pesquisa da realidade vivenciada pelos indivíduos ou grupos de estudo.

Dentre as desvantagens, o pesquisador pode conduzir as questões da pesquisa de forma a induzir, mesmo que inconscientemente, a um certo tipo de resposta, apoiando uma conclusão pretendida e tendenciosa. Ou seja, o pesquisador pode assumir interpretações que melhor se encaixem com a sua conclusão pretendida no momento de coletar e interpretar os dados da pesquisa. Os indivíduos ou grupos escolhidos podem ou não representar adequadamente a população de interesse e, como as respostas são frequentemente abertas à interpretação (subjetividade), dificulta-se a análise dos resultados com precisão.

Além disso, se os indivíduos não se sentirem à vontade com o pesquisador (não desenvolverem a confiança), pode-se não obter seus verdadeiros pontos de vista e sentimentos (os sujeitos poderão responder de uma forma que agrade mais ao pesquisador ou aos padrões socialmente aceitáveis). Ao exigir um trabalho detalhado, em profundidade, implica em investimento de tempo. Portanto, a pesquisa qualitativa denota tempo considerável para a sua execução. A pesquisa qualitativa só pode ser desenvolvida por entrevistadores bem treinados, porque não se restringe à aplicação de questionários,

mas depende da capacidade do pesquisador em campo. Por fim, a pesquisa qualitativa limita-se pelo fato de não ser possível trabalhar com amostras muito grandes, sendo necessário restringir o tamanho da amostra, o que dificulta ou impossibilita generalizações.

Redação da pesquisa qualitativa

Existem guias de redação desenvolvidos para pesquisa qualitativa, como *Standards for Reporting Qualitative Research* (SRQR) (www.equator-network.org/reporting-guidelines/srqr), que contém 21 itens; *Enhancing Transparency in Reporting the Synthesis of Qualitative Research* (ENTREQ) (www.equator-network.org/reporting-guidelines/entreq), que também possui 21 itens e; *Consolidated Criteria for Reporting Qualitative Research* (COREQ) (www.equator-network.org/reporting-guidelines/coreq), um guia mais completo, com 32 itens distribuídos em três domínios para redação. Lembre-se que estes guias, além de serem úteis para redação do projeto de pesquisa e do manuscrito da pesquisa qualitativa, também são indicados para análise crítica deste tipo de delineamento.

Análise crítica de pesquisa qualitativa

Existem algumas questões norteadoras para avaliar criticamente um artigo de uma pesquisa qualitativa:

1. o artigo descreve um problema clínico importante, abordado por meio de uma questão claramente formulada? O artigo deve descrever por que a pesquisa foi feita e qual a questão específica abordada. Não há nenhum valor científico em entrevistar ou observar pessoas simplesmente por fazê-lo. A natureza iterativa da pesquisa qualitativa permite que a pergunta de pesquisa definitiva possa não estar claramente enfatizada no começo do estudo, mas certamente deverá estar formulada quando o manuscrito for escrito;
2. foi apropriado fazer uma abordagem qualitativa? Se o objetivo da pesquisa é examinar, interpretar ou alcançar uma compreensão mais profunda de um aspecto clínico específico, provavelmente os métodos qualitativos são os mais adequados. Porém, se a pesquisa objetiva medir

ou comparar algum parâmetro quantificável, os métodos qualitativos são inapropriados;

3. como foram selecionados a situação e os sujeitos pesquisados? Na pesquisa qualitativa, pretende-se alcançar uma compreensão profunda da experiência de indivíduos ou de grupos particulares e, assim, deve-se buscar deliberadamente indivíduos ou grupos que se ajustem a este propósito. Cuidado com pesquisas qualitativas em que a amostra foi ou parece ter sido selecionada com base na conveniência, pois pode não representar os indivíduos do grupo de interesse e a saturação dos dados não ser adequada;

4. qual era a perspectiva do pesquisador? Foi levada em conta? A pesquisa qualitativa é baseada na experiência da vida real. Porém, não deve ser desprezada quando os pesquisadores declaram uma perspectiva cultural em particular ou um envolvimento pessoal com os sujeitos da pesquisa. Pelo contrário. Não existe forma de eliminar ou controlar completamente o viés do observador na pesquisa qualitativa, principalmente quando a observação participante for usada. Mas também é verdade para outras formas de coleta e análise dos dados. Dificilmente as entrevistas serão conduzidas por alguém sem nenhum ponto de vista e nenhuma perspectiva ideológica ou cultural. Por isso, é importante o pesquisador descrever a sua origem, de modo que os resultados possam ser interpretados de acordo com a perspectiva mencionada. Geralmente, os pesquisadores qualitativos preferem elaborar o manuscrito na primeira pessoa do singular, pois isto deixa explícito o papel e a influência do pesquisador;

5. quais métodos o pesquisador usou para coletar os dados? Estão descritos com detalhes suficientes? A metodologia de uma pesquisa qualitativa dificilmente poderá ser escrita de forma resumida nem ser substituída pela referência às técnicas de pesquisa de outra autoria. Provavelmente será longa e discursiva, pois está contando uma história única e sem a qual os resultados não podem ser interpretados. Assim como a estratégia de amostragem, não há regras

sobre quais detalhes devem ser incluídos. Questione se as informações disponíveis são suficientes sobre os métodos usados e se esses métodos são um modo razoável e adequado para responder à pergunta de pesquisa;

6. quais métodos o pesquisador usou para analisar os dados e quais medidas de controle de qualidade foram implementadas? A seção de análise de dados de um artigo de pesquisa qualitativa é a oportunidade para que o(s) pesquisador(es) demonstre(m) a diferença entre sentido e falta de sentido. Não é suficiente simplesmente procurar "citações interessantes" que apoiem uma teoria específica. O pesquisador deve encontrar um modo sistemático de analisar os dados e, de modo específico, procurar detectar e interpretar itens dos dados que pareçam contradizer ou desafiar as teorias resultantes da maioria. A maneira mais comum de analisar os dados qualitativos é a análise temática: o(s) pesquisador(es) vasculha(m) os textos, elabora(m) uma lista de temas amplos e aloca(m) categorias de codificação a cada um, gerando um mapa temático da análise.

Frequentemente, a análise temática é desenvolvida pelo desenho de uma matriz ou estrutura como uma coluna para cada tema de uma nova linha para cada "caso" (por exemplo, a transcrição de uma entrevista) e recorta-se e cola-se os segmentos relevantes do texto em cada célula. Outro tipo de análise temática é o método de comparação constante, no qual cada fragmento novo de dado é comparado com o resumo emergente de todos os itens anteriores, permitindo um refinamento passo a passo de uma teoria emergente.

É comum que a análise de dados qualitativos seja feita com auxílio de programas como ATLAS-TI ou NVIVO, que facilitam muito o manejo de grandes conjuntos de dados. As citações feitas por todos os entrevistados sobre um tópico específico são comparadas umas com as outras e podem ser feitas comparações específicas. Mas, um programa qualitativo de computador não analisa os dados por piloto automático. Se inserir material de má qualidade, o resultado será de má qualidade (acrônimo GIGO, *Garbage In Garbage Out*). O método mais recomendado para analisar os dados é

o método VLDRT (*Very Large Dining Room Table*), que é um método manual de comparação onde os formulários como entrevistas são marcados com caneta hidrocor.

E como demonstrar que o controle de qualidade foi alcançado? A qualidade tem mais a ver com o nível de diálogo crítico entre os pesquisadores e com o modo como as discordâncias foram expostas e resolvidas, do que com a confiabilidade interexaminadores. Lembre-se que podem haver ambiguidades inerentes aos dados;

7. os resultados possuem credibilidade? Se tiverem, são clinicamente importantes? Não podemos avaliar a credibilidade dos resultados qualitativos pela precisão e exatidão de dispositivos de mensuração, nem sua significância pelos intervalos de confiança e número necessário para tratar. O instrumento mais importante para determinar se os resultados são razoáveis e dignos de credibilidade é o simples bom senso. Um aspecto importante é verificar se os autores citam dados reais na seção de resultados, por exemplo, pela inclusão de segmentos, citações e exemplos trazidos pelos sujeitos pesquisados;

8. quais conclusões foram tiradas? São justificadas pelos resultados? Na pesquisa qualitativa, os resultados são uma interpretação dos dados. Portanto, é necessário, ao avaliar a validade da pesquisa qualitativa, perguntar se a interpretação dos dados está de acordo com o bom senso e que a perspectiva pessoal, profissional e cultural do pesquisador esteja explicitada, de modo que o leitor possa acessar a "lente" pela qual o pesquisador desenvolveu o trabalho de campo, a análise e a interpretação dos dados. Este pode ser um aspecto difícil, pois a linguagem que usamos para descrever algo tende a implicar significados e motivos que os sujeitos em si podem não compartilhar. Questione o quão bem a análise explica o porquê da maneira das pessoas se comportarem, quão compreensível esta explicação seria para um participante nesta situação e o quanto esta explicação é coerente com o que já sabemos.;

9. os achados do estudo são aplicáveis a outras situações? Uma das principais limitações da pesquisa qualitativa é

que os achados de qualquer estudo qualitativo se aplicam somente à situação limitada na qual foram obtidos. Contudo, ao promover ajustes iterativos à estrutura da amostragem à medida que o estudo se desdobra (foco progressivo), maior a probabilidade de os dados serem transferíveis para além da amostra em si, ou seja, de serem extrapolados para outros cenários e grupos.

Referências

Fontanella et al. Amostragem em pesquisas qualitativas: proposta de procedimentos para constatar saturação teórica. *Cad. Saúde Pública*. 2011;27(2):389-394.

Greenhalgh, T. Como ler artigos científicos: Fundamentos da medicina baseada em evidências. Tradução e revisão técnica: Ananyr Porto Fajardo. 5ª ed. Porto Alegre: Artmed, 2015. 376 p.

Karine de Souza, L. Pesquisa com análise qualitativa de dados: conhecendo a Análise Temática. *Arquivos Brasileiros de Psicologia*. 2019;71(2):51-67.

Pope, C; Van Royen, P; Baker, R. Qualitative methods in research on healthcare quality. *Qual Saf Health Care* 2002;11:148–152.

Victoria, C.G.; Knauth, D.R.; Hassen, M.N.A. Pesquisa qualitativa em saúde: uma introdução ao tema. Porto Alegre: Tomo Editorial, 2000. 136 p.

ISBN: 978-65-86232-63-9

CD



9 786586 232639



Faculdade de
Medicina
Universidade Federal
do Rio Grande do Sul
120 anos



PPG em Ciências da Saúde:
Ginecologia e Obstetria

HOSPITAL DE
CLÍNICAS
PORTO ALEGRE RS



fundmed
Pesquisa, Ensino e Inovação