

ISBNp
978-989-99426-5-3

Escola Superior de
Enfermagem de Coimbra

[EDITOR]

Unidade de Investigação
em Ciências da Saúde:
Enfermagem

[COORDENAÇÃO EDITORIAL]

SÍNTESE DA EVIDÊNCIA NO CONTEXTO DA TRANSLAÇÃO DA CIÊNCIA

2017



HEALTH SCIENCES
RESEARCH UNIT
NURSING

UNIDADE DE INVESTIGAÇÃO
EM CIÊNCIAS DA SAÚDE
ENFERMAGEM



**Escola Superior de
Enfermagem de Coimbra**

Editor: Escola Superior de Enfermagem de Coimbra (ESEnfC)

Coordenação editorial: Unidade de Investigação em Ciências da Saúde: Enfermagem (UICISA: E)

ISBNp: 978-989-99426-5-3

COMO SE CITA A OBRA (Normas APA 6ª edição)

Apóstolo, J. (2017). *Síntese da evidência no contexto da translação da ciência*. Coimbra, Portugal: Escola Superior de Enfermagem de Coimbra (ESEnfC).

O conteúdo científico é da responsabilidade dos autores.

ISBNp
978-989-99426-5-3

Escola Superior de
Enfermagem de Coimbra

[EDITOR]

Unidade de Investigação
em Ciências da Saúde:
Enfermagem

[COORDENAÇÃO EDITORIAL]

SÍNTESE DA EVIDÊNCIA NO CONTEXTO DA TRANSLAÇÃO DA CIÊNCIA

AUTOR DA OBRA
João Luís Alves Apóstolo

2017

FICHA TÉCNICA

TÍTULO

Síntese da evidência no contexto da translação da ciência

EDITOR

Escola Superior de Enfermagem de Coimbra

COORDENAÇÃO EDITORIAL

Unidade de investigação em Ciências da Saúde: Enfermagem

AUTORES

João Luís Alves Apóstolo

ISBNp

978-989-99426-5-3

DEPÓSITO LEGAL

(a informar pela gráfica)

COPYRIGHT

© 2017 Escola Superior de Enfermagem de Coimbra (ESEnfC)

REVISÃO FINAL

Cristina Louçano, *Lic. em Línguas e Literaturas Modernas, variante de Francês/Inglês*

Daniela Filipa Batista Cardoso, *RN, Bolseira da Unidade de Investigação em Ciências da Saúde: Enfermagem*

Maria Lucília Cardoso, *MS. em Sociologia – Bolseira de Investigação da UICISA: E, Escola Superior de Enfermagem de Coimbra*

MAQUETIZAÇÃO

Eurico Nogueira, *MS em Tecnologias de Informação Visual*

REVISÃO DOCUMENTAL

Centro de Documentação e Informação da ESEnfC

APOIO TÉCNICO

Cristina Louçano, *Secretariado da Unidade de Investigação em Ciências da Saúde: Enfermagem*

GRÁFICA

Tipografia Lousanense, Lda – Lousã

TIRAGEM

400 exemplares

ANO DE PUBLICAÇÃO

2017

ÍNDICE

NOTA CURRICULAR.....	1
NOTA DO EDITOR.....	3
<i>Manuel Alves Rodrigues</i>	
INTRODUÇÃO.....	5
CAPÍTULO I	
O CICLO DE TRANSLAÇÃO DA CIÊNCIA.....	9
1.1 A geração, síntese, transferência e implementação de evidências para os cuidados de saúde.....	11
1.1.1 Barreiras e lacunas no processo de translação da ciência.....	13
1.2 Estruturas específicas para a síntese, transferência e implementação da evidência	16
CAPÍTULO II	
SÍNTESE DA EVIDÊNCIA.....	21
2.1 Minimizar o risco de viés no processo de síntese da evidência.....	23
2.1.1 Minimizar o risco de viés numa revisão quantitativa.....	25
2.1.1.1 Instrumentos de avaliação crítica.....	28
2.1.1.1.1 Instrumento de Avaliação Crítica para Estudos Randomizados Controlados (ERC; JBI, 2016).....	29
2.1.1.1.2 Instrumento de Avaliação Crítica para Estudos Quasi-experimentais – estudos experimentais sem alocação randomizada (JBI, 2016).....	35
2.1.1.1.3 Instrumento de Avaliação Crítica de Estudos de Coorte (JBI, 2016).....	38
2.1.1.1.4 Instrumento de Avaliação Crítica para Estudos Caso-Controllo (JBI, 2016).....	42
2.1.1.1.5 Instrumento de Avaliação Crítica de Estudos de Caso (JBI, 2016).....	45
2.1.1.1.6 Instrumento de Avaliação Crítica de Série de Casos (JBI, 2016).....	46
2.1.1.1.7 Instrumento de Avaliação Crítica de Estudos de Prevalência (Munn, Moola, Lisy, Rittano, & Tufanaru, 2015).....	49
2.1.1.1.8 Instrumento de avaliação crítica para estudos de precisão diagnóstica (JBI, 2015b; Whiting et al., 2011).....	52
2.1.1.1.9 Instrumento de Avaliação Crítica para Estudos Transversais Analíticos (JBI, 2016).....	55

2.1.2	Minimizar o risco de viés numa revisão qualitativa.....	57
2.1.2.1	Avaliação crítica e validade dos estudos qualitativos.....	60
2.1.2.1.1	Instrumento de avaliação crítica para estudos qualitativos (JBI, 2014a; Lockwood, Munn, & Porritt, 2015).....	60
2.1.2.1.2	Instrumento de Avaliação Crítica para Estudos de Texto e Opinião (McArthur, Klugarova, Yan, & Florescu, 2015).....	64
2.1.3	Minimizar o risco de viés numa revisão de revisões.....	65
2.1.3.1	Instrumento de Avaliação Crítica para Revisões Sistemáticas e Síntese de Investigação (Aromataris et al., 2015; JBI, 2014d, pp. 30-32).....	65

CAPÍTULO III

TIPOS DE REVISÃO SISTEMÁTICA DA LITERATURA.....71

3.1	RSL de evidência de eficácia.....	72
3.1.1	Meta-análise.....	75
3.2	RSL de evidência de significado.....	83
3.3	RSL de prevalência e incidência.....	87
3.4	RSL de associação (etiologia).....	89
3.5	RSL de precisão de testes de diagnóstico.....	92
3.6	RSL de métodos mistos (<i>Mixed Methods</i>).....	94
3.7	RSL de revisões de revisões (<i>Umbrella</i>).....	99
3.8	Revisão <i>scoping</i>	101

CAPÍTULO IV

RESULTADOS DE AVALIAÇÃO CRÍTICA E DISCUSSÃO.....107

CONSIDERAÇÕES FINAIS.....115

ÍNDICE DE TABELAS

Tabela 1. Avaliação da qualidade metodológica dos ERCs incluídos.....	109
Tabela 2. Avaliação da qualidade metodológica dos estudos qualitativos incluídos.....	111
Tabela 3. Avaliação da qualidade metodológica das revisões incluídas.....	112

ÍNDICE DE FIGURAS

Figura 1. Fluxograma de decisão para a seleção do modelo estatístico de meta-análise. Adaptado de “Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness” por Tufanaru, Munn, Stephenson, e Aromataris, 2015, <i>International Journal of Evidence-Based Healthcare</i> , 13(3), p. 203.....	78
Figura 2. Gráfico de floresta comparando a eficácia da água da torneira versus a solução salina estéril sobre as taxas de infecção de feridas agudas e crônicas. Adaptado de “The effectiveness of cleansing solutions for wound treatment: A systematic review” por P. Queirós, E. Santos, J. Apóstolo, D. Cardoso, M. Cunha, e M. Rodrigues, 2014, <i>JBI Database of Systematic Reviews and Implementation Reports</i> , 12(10), p. 133.	82

SIGLAS

AHRQ - Agency for Healthcare Research and Quality

CASP - *Critical Appraisal Skills Programme*

CDSR - *Cochrane Database of Systematic Reviews*

CENTRAL - *Cochrane Central Register of Controlled Trials*

CINAHL - *Cumulative Index to Nursing and Allied Health Literature*

CMR - *Cochrane Methodology Register*

CoCoPop - Condição, Contexto e População

COhNECT+ - *Clinical Online Network of Evidence for Care and Therapeutics +*

CSBE - Cuidados de Saúde Baseados em Evidência

CSRTP - *Comprehensive Systematic Review Training Program*

DARE - *Database of Abstracts of Reviews of Effects*

EMBASE - *Excerpta Medica Database*

ERIC - *Education Resources Information Center*

GRADE - *Grading of Recommendations Assessment, Development and Evaluation*

GRC - Grupos de Revisão Cochrane

IC - Intervalo de confiança

ITT - *Intention-to-treat*

JBI - Joanna Briggs Institute

MEDLINE - *Medical Literature Analysis and Retrieval System Online*

MM - Métodos Mistos

NOCs - Normas de Orientação Clínica

OR - Odds Ratio

PBE - Prática Baseada em Evidência

PCC - População, Conceito e Contexto

PEDro - *Physiotherapy Evidence Database*

PEO - População, Exposição de interesse e *Outcome*

PICo - População, fenómeno de Interesse e Contexto

PICO - População, Intervenção, Comparador e *Outcome*

PIRD - População, teste diagnóstico de Interesse, teste de Referência, Diagnóstico de interesse

QUADAS - *Quality Assessment of Diagnostic Accuracy Studies*

ROC - *Receiver Operating Characteristic*

RR - Risco Relativo

RS - *Revisão Scoping*

RSL - *Revisão Sistemática da Literatura*

SCOPUS - Bibliographic database owned by Elsevier and covering the fields of Science, Technology, Medicine, social sciences, and Arts & Humanities

SMD - *Standardized Mean Difference*

WMD - *Weighted Mean Difference*

NOTA CURRICULAR



João Luís Alves Apóstolo nascido em Portugal em 1964. Agregação em Enfermagem (2017), Pós-doutoramento em Ciências de Enfermagem (2013), Doutoramento em Ciências de Enfermagem (2008), Mestrado em Toxicodependências e Patologias Psicossociais (2000), Graduação em Enfermagem (1988), Especialização em Saúde Mental e Enfermagem Psiquiátrica (1997), Pós-Graduação em Administração de Serviços de Saúde (2000). Enfermeiro, generalista e especialista, de 1988 a 2000, nos Hospitais da Universidade de Coimbra, Centro Hospitalar de Coimbra e Hospital Psiquiátrico Sobral Cid, Portugal. Desde 2001 é professor na Escola Superior de Enfermagem de Coimbra, Portugal nas áreas de metodologias de investigação, cuidados de saúde baseada em evidências e enfermagem de saúde do idoso. Em 2013, acede à categoria de Professor Coordenador. Investigador da Unidade de Investigação em Ciências da Saúde: Enfermagem em projetos de investigação relacionados com saúde mental e envelhecimento. Diretor adjunto do Portugal Centre for Evidence-based Practice: A JBI Centre of Excellence – PCEBP. Membro do “Cognitive Decline Group” European Innovation Partnership on Active and Healthy Ageing (EIP-AHA). Revisor por pares para 13 revistas científicas internacionais. Autor em 90 artigos publicados em revistas com revisão por pares; 70 resumos em Atas de eventos científicos; 8 capítulos de livros e 2 livros, de edição nacional e internacional.

Curriculum DeGóis – Endereço:

<http://www.degois.pt/visualizador/curriculum.jsp?key=9822639125721612>

NOTA DO EDITOR

Em 2010, após a segunda avaliação e acreditação da Unidade de Investigação em Ciências da Saúde: Enfermagem (UICISA: E) pela Fundação para a Ciência e Tecnologia (FCT), reconhecia-se a importância da aquisição de conhecimento sobre Prática Baseada na Evidência (PBE), com particular foco na síntese da evidência. Motivados pelo Professor Alan Pearson, e suportado por uma equipa de investigadores que realizou em Inglaterra a formação Comprehensive Systematic Review Training Program (CS RTP), foi possível submeter um plano de ação para 3 anos para a constituição de um centro do Joanna Briggs Institute (JBI). Este plano foi aprovado dando lugar à criação do Portugal Centre for Evidence Based Practice: an affiliate centre of the JBI (PCEBP) que passa a fazer parte de uma relevante rede internacional de mais de 70 centros afiliados e colaboradores. Inaugurava-se assim o Eixo Estratégico de Desenvolvimento para a Síntese da Ciência da UICISA: E.

O autor deste trabalho, docente da Escola Superior de Enfermagem de Coimbra (ESEnfC) e investigador da UICISA: E, pela sua dedicação e aprendizagem continuada, no contexto desta rede colaboradora, soube acompanhar, ao longo do tempo, a evolução dos conhecimentos sobre a ciência da síntese. Deste modo, pôde agora organizar de forma valiosa várias abordagens e técnicas diferenciadas no contexto do ciclo de translação da ciência (geração, síntese, transferência e implementação da ciência).

O valor científico, técnico e didático deste documento é observável ao longo dos diferentes capítulos. No capítulo 2, o autor analisa com precisão as questões relativas à minimização do erro, seja em revisão quantitativa seja em revisão qualitativa. Salienta a importância de minimizar o erro e de assegurar que a evidência sintetizada, a partir do universo do conhecimento disponível, seja a que melhor responde às necessidades identificadas e às questões levantadas. No capítulo 3, aborda os diversos tipos de Revisão sistemática da literatura, iniciando com a problemática de revisão sistemática da literatura (RSL) de evidência de eficácia.

A discussão no contexto da rede JBI, ao longo destes anos, tem feito emergir diferentes tipologias de RSL, atualmente mais de uma dezena. O autor oferece, de forma atualizada e organizada, algumas das técnicas mais relevantes.

Esta obra organiza a experiência do autor construída, até ao momento, através: de ações de formação, acreditadas pelo JBI, especificamente o curso Comprehensive Systematic Review Training Program (CSRTP), ministradas enquanto formador *Train-the-Trainer*; do estudo dos modelos e tecnologias JBI; e da otimização da produtividade que conduziu ao reconhecimento do PCEBP como Centro de Excelência JBI.

Na condição de Coordenador Científico da Unidade de Investigação, não somente agradeço o empenho do autor no percurso de construção e sustentação das atividades de síntese da ciência como também felicito-o por este contributo que serviu de base às suas provas académicas de agregação em Enfermagem. Este contributo certamente será utilizado de forma muito útil na atividade científica, pedagógica e clínica.

Fortalecidos por este produtivo percurso que é preciso continuar, abrimo-nos agora ao novo desafio de capacitar uma equipa na UICISA: E para a Implementação da evidência. Através de futuros cursos, nomeadamente o curso JBI Evidence-Based Clinical Fellowship Program (EBCFP), almeja-se consolidar o ciclo completo da translação da ciência. Pretende-se envolver cada vez mais investigadores, académicos, clínicos mas também cidadãos, a quem se dirige, em última estância, a prática informada pela melhor evidência.

Manuel Alves Rodrigues

Professor Coordenador Principal
Coordenador Científico da UICISA: E

INTRODUÇÃO

O conceito *Evidence-Based* (baseado em evidência) tem tido rápido desenvolvimento desde que surgiu no início de 1990, inicialmente sob a liderança pelo professor David Sackett da Universidade de Oxford (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Evoluiu para além do seu uso inicial no campo da medicina e da prática médica, incluindo outras facetas e profissionais e fazendo emergir os conceitos de prática baseada em evidência (PBE) e de cuidados de saúde baseados em evidência (CSBE; Dawes et al., 2005; Pearson, Jordan, & Munn, 2012).

A melhor prática é uma característica central de todos os serviços de saúde com base em evidências. A importância de basear as políticas de saúde e práticas de cuidados de saúde sobre a melhor evidência internacional disponível e de traduzir o conhecimento ou evidência em ação - investigação translacional - tem vindo a ser cada vez mais enfatizada em diversos sectores da saúde e em diversos países (Pearson et al., 2012). CSBE é um processo que identifica políticas ou perguntas clínicas dando-lhes respostas através da geração de conhecimento e evidência para orientar a prestação de cuidados de saúde de forma eficaz e adequada, mais realista e significativa para populações, culturas e contextos específicos.

Subjacente a estes conceitos está a tomada de decisão clínica que considera a melhor evidência disponível, o contexto no qual o cuidado é prestado, a preferência do utente, e o julgamento clínico do profissional de saúde (Pearson, Wiechula, Court, & Lockwood, 2005). Considerando que, para além da melhor evidência disponível, também o julgamento do clínico e as preferências do utente devem ser consideradas na tomada de decisão clínica, muito recentemente, tem vindo a ser discutido o conceito de prática informada pela evidência. Como esta é uma discussão em aberto, que teremos a oportunidade de retomar adiante, para os devidos efeitos, no contexto deste documento, utilizaremos o conceito de prática baseada em evidência (PBE) por ser internacionalmente aceite e disseminado.

No contexto da PBE e dos CSBE, a evidência produzida é criticamente avaliada, sintetizada e transferida para os contextos da prática para ser utilizada pelos profissionais de saúde que, em seguida, avaliam o seu impacto sobre os resultados de saúde, sistemas de saúde e prática profissional. Como preconizado pelos diversos modelos da PBE, este é um processo cíclico que vai da produção e síntese crítica do conhecimento à avaliação do impacto da sua implementação na prática (Pearson et al., 2005; Pearson et al., 2012; Schaffer, Sandau, & Diedrick, 2013; Titler, 2008).

A PBE acrescenta elevado valor aos cuidados de saúde, melhora a experiência do utente (incluindo a qualidade e confiabilidade dos cuidados de saúde) e os resultados em saúde, e reduz os custos (Melnik, Gallagher-Ford, Long, & Fineout-Overholt, 2014). A adoção da PBE tem potencial para melhorar a qualidade dos cuidados e os resultados nos utentes. Assim, a adoção, implementação e sustentabilidade da PBE torna-se cada vez mais importante para a saúde, para as organizações de saúde e para os profissionais (Aarons, Ehrhart, & Farahnak, 2014; Wise, 2009).

Além disso, as orientações internacionais e nacionais, tais como as que constam no Plano Nacional de Saúde Português 2012-2016 (Direção Geral da Saúde, 2012), e nos documentos do Conselho Nacional de Enfermagem (Ordem dos Enfermeiros, 2006, 2012), do Conselho Internacional de Enfermeiros (International Council of Nurses, 2012), e da Agency for Healthcare Research and Quality (AHRQ, 2001) recomendam a implementação da PBE argumentando que: a tomada de decisão é simplificada; a incerteza, o risco e a variabilidade são reduzidos; e a qualidade dos cuidados é melhorada.

Como o objetivo dos CSBE é o melhor resultado em saúde, a utilização da melhor prática disponível é a chave para o sucesso. Esta componente tem três elementos: mudar a prática clínica, incorporar evidência através da mudança do sistema/organização e avaliar o impacto da utilização de evidências no sistema de saúde, no processo de cuidados e nos resultados em saúde (Joanna Briggs Institute - JBI, 2013a; Pearson et al., 2012; Schaffer et al., 2013).

As crenças no valor da PBE, a capacidade para implementar as melhores práticas, os contextos ou ambientes facilitadores da implementação da PBE e o apoio administrativo são alguns dos fatores facilitadores referenciados para a adoção da PBE (Melnik et al., 2014). Porém, a adoção, implementação e sustentabilidade da PBE no contexto da saúde, das organizações de saúde e dos prestadores de cuidados de saúde, em que os enfermeiros são elementos centrais, coloca uma série de desafios que contribuem para as muitas lacunas ainda existentes entre investigação, prática e política de saúde (Pearson et al., 2012).

Neste processo cíclico, a síntese da evidência é uma peça crítica que carece de peritagem nesta área, e por essa razão, merecerá especial atenção neste documento. Os profissionais de saúde, consumidores, investigadores e enunciadores de políticas

são inundados com quantidades incontroláveis de informação. É improvável que todos tenham o tempo, as competências e os recursos para procurar, avaliar e interpretar essa evidência e incorporá-la nas decisões em cuidados de saúde de maneira eficaz, viável, e significativa para populações específicas, culturas e contextos. Os clínicos ocupados, raramente, conseguem manter-se a par de todos os desenvolvimentos atuais na sua área e muito menos têm a oportunidade de participar no processo de realização de uma revisão sistemática (Pearson et al., 2014).

Assim, algumas estruturas mundiais têm vindo a desenvolver metodologias para suportar o desenvolvimento da síntese e implementação da evidência. Entre outras, a Colaboração Cochrane e o JBI destacam-se pela dimensão da sua estrutura mundial. Não obstante o reconhecimento internacional da Colaboração Cochrane, pela ligação que temos à JBI e ao Portugal Centre for Evidence Based Practice (PCEBP): a JBI Centre of Excellence adotamos neste documento, maioritariamente, a visão, o modelo e a metodologia proposta pelo JBI.

O documento que aqui se apresenta foi desenvolvido no âmbito da lição “Síntese da evidência no contexto da translação da ciência” para prestação de provas para habilitação ao título académico de agregado em Ciências de Enfermagem, no Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto. Foram atualizados os instrumentos de avaliação crítica da JBI entretanto disponibilizados, e consultados no dia 7 de julho de 2017, em <http://joannabriggs.org/research/critical-appraisal-tools.html>.

Este documento está organizado em quatro pontos principais. No ponto 1 apresenta-se, genericamente, o ciclo de translação da ciência desde a geração de evidências para os cuidados de saúde, à síntese crítica, à transferência e à implementação da evidência na prática, considerando ainda as estruturas mundiais reconhecidas neste processo cíclico.

Apresenta-se no ponto 2, com particular destaque, o processo de síntese da evidência. Inicia-se com considerações sobre a minimização do risco de viés no processo de síntese da evidência quantitativa, qualitativa e de revisões de revisões.

O ponto 3 refere-se à descrição e exemplificação das especificidades das diferentes tipologias de revisão sistemática da literatura (RSL), entre as quais as de eficácia, significado (habitualmente designadas por revisões qualitativas), prevalência e incidência, associação (etiologia), precisão de testes de diagnóstico, métodos mistos, revisão de revisões e revisões de abrangência (a que nos referiremos com o nome de revisão *scoping* - RS - traduzido a partir do inglês *scoping review*). As referidas especificidades e exemplos são relativos ao título, questão de revisão, elementos a considerar em relação ao PICO (população, intervenção, comparador e resultados) ou respetiva adaptação para cada tipologia de revisão, e a metodologia de síntese de dados.

Finaliza-se, no ponto 4, com a apresentação de resultados e discussão sobre os processos de avaliação crítica de revisões desenvolvidas e de outras fragilidades verificadas habitualmente nas RSLs relacionadas com a avaliação crítica, com os níveis de evidência e graus de recomendação para a prática.

CAPÍTULO I

O CICLO DE TRANSLAÇÃO DA CIÊNCIA

A implementação do conhecimento na prática dos cuidados e a avaliação do seu impacto deverá ser o fim último de um processo cíclico e complexo. A visão da ciência da translação enfatiza a aplicação dos resultados das ciências básicas ou laboratoriais na prática clínica através da realização de ensaios clínicos. Contudo a translação da ciência é um processo mais amplo do que a realização de ensaios clínicos para testar descobertas. Inicia-se com a translação das perguntas que surgem da necessidade de conhecimento no mundo real para a investigação e finaliza com a translação dos resultados da investigação em ação clínica ou política (Pearson et al., 2012).

Este processo, não obstante as reconhecidas falhas e barreiras, tal como nos referiremos mais adiante, inclui a investigação de métodos, intervenções e variáveis que influenciam a adoção da PBE pelos indivíduos e pelas organizações para melhorar a tomada de decisão clínica e operacionalizar os cuidados de saúde. Deve ser orientado por um modelo conceptual que organiza as estratégias para a implementação, clarifica as variáveis externas que podem influenciar a adoção da PBE e constrói uma base de conhecimento científico para este campo de investigação (Titler & Everett, 2001).

Há uma série de modelos que tentam representar os componentes dos CSBE. Com objetivo de apresentar uma visão sucinta dos modelos que suportam a adoção da PBE que podem ser aplicados na prática dos CSBE e de avaliar a sua utilidade em ambientes clínicos e educacionais, Schaffer et al. (2013) identificaram os seguintes seis modelos mais usuais: (1) *ACE star model of knowledge transformation*; (2) *advancing research and clinical practice through close collaboration*; (3) *Iowa model*; (4) *Johns Hopkins nursing evidence based practice model*; (5) *promoting action on research implementation in health services framework*; e (6) *Stetler model*. Os autores descrevem as principais características de cada um dos referidos modelos e fornecem uma avaliação do modelo de utilidade, com base em critérios específicos, focados na facilitação do processo de PBE e mudança da prática. Outros modelos referenciados na literatura são: o *five stage*

model of evidence-based healthcare, o *Agency for Healthcare Research and Quality model of knowledge transfer* (Titler, 2008), e o *JBI model of evidence-based healthcare* que traduzimos para português como modelo JBI de cuidados de saúde baseados em evidências (JBI, 2013a; Pearson et al., 2012).

Os diferentes modelos descritos propõem que a implementação/translação da ciência se processe em vários passos. Estes consideram, na generalidade, a produção e síntese crítica do conhecimento, a sua disseminação, a sua adoção e implementação no contexto das organizações e, finalmente, a avaliação do impacto nos utentes e organizações dessa implementação.

Neste contexto, o modelo JBI de CSBE (Pearson et al., 2005) foi construído a partir da experiência prática no campo da PBE, do emergente trabalho internacional do JBI e da sua colaboração internacional constituída por 80 centros colaboradores e grupos afiliados. Este modelo foi (re)considerado em agosto de 2015 na reunião de diretores do JBI em Adelaide, na qual colaborámos, e vertido no documento editado pela JBI em 2015 (JBI, 2015a).

Como já referimos, a PBE pode ser conceptualizada como a tomada de decisão clínica que considera a melhor evidência disponível, o contexto no qual o cuidado é prestado, a preferência do utente e o julgamento profissional do clínico. Portanto, a questão do contexto da preferência do utente e o julgamento do profissional de saúde são aspetos centrais. Nesta linha tem havido algum debate na literatura e nas reuniões científicas sobre o conceito de prática informada pela evidência.

Melnyk e Newhouse (2014), a partir de uma pesquisa da qual resultaram 327 entradas publicadas na base de dados PubMed, referem que é maioritariamente atribuído o mesmo significado aos dois conceitos. Os autores defendem que deve ser usado o conceito de PBE, argumentando que uma mudança na terminologia irá enfraquecer o movimento da PBE. No entanto, existem igualmente fortes argumentos para a necessidade de uma mudança que permita a reflexão mais precisa do papel e da utilização da evidência na prática.

No seu modelo novo, proposto em 2015, e apesar dos argumentos que apresentamos anteriormente, o JBI argumenta que é mais adequada a utilização do termo *informado* comparativamente ao termo *baseado*. Para o JBI há mais para a tomada de decisão clínica do que apenas evidências. De facto, num folheto informativo de boa prática o JBI não apresenta diretivas aos seus utilizadores, antes fornece a melhor evidência disponível para que a prática seja informada (JBI, 2015a).

Na tomada de decisões clínicas, os profissionais de saúde estão preocupados com a eficácia e viabilidade da sua abordagem e, também, com o facto da evidência e da tomada de decisão clínica ser oportuna, adequada e significativa para as pessoas ou

comunidades. A viabilidade, adequação, significado e eficácia das práticas de saúde podem ser informadas pela melhor evidência disponível, o contexto no qual o cuidado é prestado, o indivíduo, o julgamento e a experiência do profissional de saúde (JBI, 2015a).

No modelo JBI os CSBE são representados como um processo cíclico que começa com questões clínicas, preocupações ou interesses dos clínicos. A resposta a estas questões é dada através da geração de conhecimento e de dados para atender com eficácia e de forma adequada a essas necessidades de modo a que sejam aplicáveis e com significado para as populações, culturas e contextos.

O modelo JBI tem quatro principais componentes do processo que são: a geração de evidências para os cuidados de saúde, a síntese da evidência, a transferência das evidências, e a utilização das evidências na prática. A evidência produzida pelos estudos primários é avaliada no que respeita à sua qualidade, sintetizada e transferida, e disponibilizada aos profissionais e aos locais de prestação de cuidados que, em seguida, a utilizam e avaliam o seu impacto nos resultados em saúde dos indivíduos, nos sistemas de saúde e na prática profissional (JBI, 2015a; Pearson et al., 2005).

1.1 A geração, síntese, transferência e implementação de evidências para os cuidados de saúde

A geração de evidências inicia-se com a produção de investigação primária - empírica - rigorosa utilizando os mais diversos desenhos de investigação. No entanto, reconhece-se que quando não existe evidência produzida a este nível, podem-se considerar outras fontes de evidência, como a opinião de peritos, porque estas podem representar a *melhor evidência disponível* para a pergunta específica. O modelo JBI também reconhece que os profissionais de saúde, para informar a sua prática quotidiana, devem considerar para além da evidência de eficácia, outras formas de evidência, como evidência sobre aplicabilidade, adequação e significado (JBI, 2014a, 2015a; Pearson et al., 2005) que definimos de seguida:

- A evidência sobre eficácia (*effectiveness*) relaciona-se com o efeito benéfico da intervenção e com a sua segurança;
- A evidência sobre aplicabilidade (*feasibility*) está relacionada com aspetos de relação custo-eficácia, recursos físicos, culturais, e financeiros, práticas disponíveis e experiência e/ou competência clínica suficiente;
- A evidência sobre adequação ao contexto (*appropriateness*) relaciona-se com a aceitabilidade face à cultura, valores e crenças dos utentes, com a transferibilidade e aplicabilidade à população de interesse, e com a adaptabilidade a circunstâncias diversas;

- O significado para as pessoas (*meaningfulness*) refere-se à contingência da evidência estar associada a experiências positivas e não estar associada a experiências negativas.

A síntese da evidência é a organização dos resultados da investigação primária e pode ser feita com recurso a revisões sistemáticas, revisões rápidas e *clinical guidelines* (que em português traduzimos como normas de orientação clínica - NOCs), entre outras.

A revisão sistemática é o núcleo central e o padrão ouro da síntese de evidências, que é por si uma metodologia de investigação - investigação secundária (Higgins & Green, 2011; JBI, 2014a). Como desenvolveremos à frente, a metodologia de revisão sistemática tem evoluído e neste momento há diversas tipologias de revisões que, para além das tradicionais RSL de eficácia, incluem também RSL qualitativa, de análise económica, de associação (etiologia), de precisão de diagnóstico, revisões de revisões, entre outras.

Em menor escala os resumos de evidências ou revisões rápidas, que se desenvolvem em menos de cinco semanas, comparativamente com as RSL, que podem demorar entre seis meses a dois anos, surgiram como uma abordagem simplificada para sintetizar evidências de forma mais rápida e atempada para atender às necessidades dos decisores políticos, partes interessadas e outros utilizadores do conhecimento (Khangura, Konnyu, Cushman, Grimshaw, & Moher, 2012).

As NOCs são fontes de informação resumida sobre as práticas específicas relacionadas com o atendimento ao utente para orientar a decisão clínica dos profissionais de saúde. Podem, ou não, basear-se nos resultados de uma revisão sistemática da evidência internacional.

A transferência das evidências refere-se ao ato de deslocação das evidências sintetizadas para os profissionais de saúde, unidades e sistemas de saúde a nível mundial, através de revistas científicas, outras publicações, educação e treino, e também através do suporte à decisão clínica. A transferência é mais que a divulgação ou distribuição de informação. Inclui o desenvolvimento de estratégias para identificar públicos-alvo, como clínicos, gestores, formuladores/enunciadores de políticas, consumidores, etc. Inclui ainda o desenho, implementação e métodos para organizar e transferir a informação que é compreensível e utilizável na tomada de decisões. Neste processo, é fundamental que a informação seja compreensível, ajustada ao contexto e às necessidades de informação do público-alvo. A informação deve também poder ser integrada em sistemas de informação, como sistemas de apoio à decisão clínica.

Para que a evidência se torne de fácil acesso e consumo para os clínicos e os utentes, a informação resultante das RSLs pode ser resumida e condensada numa variedade de formatos. Por exemplo, o JBI através da sua base de dados JBI CONNECT+ (*Clinical Online*

Network of Evidence for Care and Therapeutics), disponível através do site <http://connect.jbiconnectplus.org/Default.aspx>, fornece, para além das RSL, a seguinte informação resumida:

- Fichas de informação da melhor prática - são pequenos resumos com base nos resultados e recomendações de RSLs que permitem fácil disseminação e o acesso dos profissionais de saúde às principais questões e recomendações que foram recolhidas a partir de um grande volume de material;
- Folhas de informação ao consumidor - são resumos padronizados sobre uma ampla gama de intervenções de cuidados de saúde e atividades dirigidas aos consumidores de cuidados de saúde, ou seja, utentes, familiares e cuidadores. Cada ficha de informação do consumidor é baseada na melhor evidência internacional disponível e atualizada anualmente no caso de existir nova informação;
- Práticas recomendadas baseadas em evidências - são descrições de intervenções ou procedimentos que descrevem e/ou recomendam certas práticas sobre temas clínicos selecionados. As práticas recomendadas são baseadas nas melhores evidências disponíveis;
- Sumários de evidência - são resumos curtos que sumariam a evidência internacional existente sobre intervenções/orientações em cuidados de saúde comuns e atividades, baseando-se em pesquisas estruturadas da literatura.

A implementação, no contexto do modelo JBI, é definida como um conjunto de atividades de capacitação destinadas a envolver os principais interessados com as evidências, para que a tomada de decisões possa ser informada e gerar melhoria sustentada na qualidade da prestação de cuidados de saúde (JBI, 2014a, 2015a; Pearson et al., 2005). Considerando que, no conjunto de estudos incluídos numa RSL, pode existir heterogeneidade clínica, geográfica e cultural, é discutível a aplicabilidade prática dos seus resultados gerados por essa revisão em contextos diferentes (West et al., 2010). Por essa razão, os processos de implementação devem incluir uma análise situacional, a facilitação da mudança prática e avaliação do processo e do resultado.

A implementação da evidência é um processo complexo e não cabe no âmbito deste trabalho o seu desenvolvimento. No entanto, desenvolveremos algum pensamento sobre algumas barreiras e lacunas existentes e estratégias para facilitar este processo.

1.11 Barreiras e lacunas no processo de translação da ciência.

A tradução dos resultados da investigação em melhorias sustentáveis nos resultados clínicos e nos resultados dos utentes continua a ser um obstáculo substancial para a

melhoria da qualidade do atendimento. A adoção da PBE não é o padrão dos cuidados de saúde no mundo (Melnyk et al., 2014). Existe uma lacuna entre a investigação e a prática ou a política (Oliver, Innvar, Lorenc, Woodman, & Thomas, 2014). Este fosso tem várias dimensões como lacunas e barreiras que limitam a implementação da evidência na prática podendo, muitas vezes, decorrer até duas décadas para que os resultados da investigação original possam torna-se parte da prática clínica de rotina (AHRQ, 2001).

Neste processo cíclico e complexo, que tem por objetivo melhorar os resultados em saúde através da tradução do conhecimento em ação, são consideradas três grandes falhas ou lacunas (Pearson et al., 2012; Pearson, Weeks, & Stern, 2011) como descrevemos a seguir:

- A primeira lacuna ocorre entre as necessidades de conhecimento identificadas pelos utentes e comunidade, pelos clínicos, pelos governos e pelas organizações, e o trabalho realizado pelos investigadores durante o processo de investigação (a lacuna entre a necessidade de conhecimento e a descoberta de novos conhecimentos).
- A segunda entre a investigação de *descoberta* realizada, seja ela teórica, epidemiológica, ou animal, e a tradução destes resultados na investigação clínica ou política (a lacuna entre a descoberta de novos conhecimentos e a aplicação clínica desse conhecimento).
- Em terceiro lugar, na investigação clínica, o impacto/ação dos resultados da investigação clínica ou política na ação ao nível clínico ou político não é avaliado consistentemente, constituindo-se como a terceira lacuna (a lacuna entre a aplicação clínica e o desenvolvimento de ações clínicas ou políticas de rotina; Pearson et al., 2011; Pearson et al., 2012).

De facto, a necessidade de melhorar a translação dos resultados básicos e fundamentais de investigação para a prática clínica de rotina foi uma das principais observações da *A review of UK health research funding* (Cooksey, 2006), considerando-se duas lacunas importantes na tradução de investigação em saúde: (i) traduzir as ideias da investigação básica e clínica no desenvolvimento de novos produtos e abordagens para o tratamento da doença; (ii) implementar esses novos produtos e abordagens para a prática clínica.

Do ponto de vista das organizações, as limitações de tempo, a cultura e a filosofia organizacional (que é “a maneira que sempre fizemos aqui”), a resistência da liderança, e o conhecimento inadequado ou treino para aceder ou avaliar criticamente a evidência, são algumas das barreiras identificadas para a implementação de PBE (Melnyk et al., 2014). Tendo em conta as barreiras e os facilitadores da PBE ao nível organizacional,

afirma-se que, além de características pessoais, a liderança e o clima organizacional são elementos fundamentais para a implementação da PBE e sua sustentabilidade (Aarons, 2006; Aarons, Hurlburt, & Horwitz, 2011; Apóstolo, Cardoso, & Rodrigues, 2016; Ehrhart, Aarons, & Farahnak, 2014; Ehrhart, Schneider & Macey, 2014; Gifford, Davies, Edwards, Griffin, & Lybanon, 2007).

As características pessoais, fundamentais para implementar com sucesso a PBE, incluem a idade, o nível de educação, a formação, a experiência profissional e os conhecimentos e atitudes face à PBE. É fundamental, que os utilizadores da evidência tenham conhecimentos de investigação e de tecnologia de informação, bem como a capacidade para interpretar a literatura e para aplicar os resultados veiculados na literatura (a evidência) aos casos individuais (o conhecimento relativamente à PBE; Aarons et al., 2011; Apóstolo et al., 2016).

Há estratégias focadas em aspetos pessoais, na liderança e no clima organizacional para facilitar o processo de implementação e sustentar a PBE com sucesso. Exemplos dessas estratégias são: (a) o desenvolvimento de conhecimentos e habilidades individuais dos clínicos e dos líderes em PBE; (b) a promoção de um contexto e cultura de apoio à PBE, incluindo a disponibilidade de recursos e de mentores em PBE; (c) o desenvolvimento de líderes em cuidados de saúde que podem liderar equipas que criam uma emocionante visão, missão e objetivos estratégicos para a implementação de todo o sistema da PBE; (d) tempo, recursos, mentores e ferramentas suficientes para que os clínicos se dediquem à PBE; (e) expectativas claras sobre o papel dos clínicos com formação em prática avançada na implementação e sustentação da PBE; e (f) o reconhecimento ou sistema de recompensa para aqueles que estão totalmente comprometidos com este esforço (Apóstolo et al., 2016; Melnyk et al., 2014).

Os serviços de saúde de sucesso do futuro serão aqueles que investirem na procura constante de informações necessárias para orientar práticas, desenvolver líderes clínicos, bem como estabelecer e manter uma cultura de melhoria na prática diária. Se o foco atual na PBE é conduzir a reduções na variabilidade das práticas em uso, bem como, a redução de custos (decorrentes da utilização do produto adequado e melhores resultados de saúde) e, ainda a uma melhoria nos resultados da saúde, é necessária uma estratégia para desenvolver capacidades nas organizações, nos líderes, nos clínicos e, em última análise, nos utentes para otimizar a probabilidade da implementação da PBE bem-sucedida (Apóstolo et al., 2016).

Por essa razão, algumas estruturas mundiais têm vindo a desenvolver essa capacidade e conhecimento metodológico, e a organizar, supervisionar e desenvolver metodologias para suportar o desenvolvimento da síntese e implementação da evidência tal como descreveremos de seguida.

1.2 Estruturas específicas para a síntese, transferência e implementação da evidência

A colaboração Cochrane e a colaboração JBI são as duas estruturas internacionais com maior impacto na síntese e disponibilização de evidência. Esta evidência pode ser pesquisada nas suas plataformas:

- Cochrane - disponível em <http://www.cochranelibrary.com>.
- JBI - disponível em <http://connect.jbiconnectplus.org> e em <http://journals.lww.com/jbisrir/Pages/issuelist.aspx>.

Para além destas, existem outras estruturas com recursos específicos que disponibilizam NOCs e recomendações para a prática, das quais destacamos a AHRQ. A mesma pode ser consultada em: <http://www.guideline.gov/index.aspx>. Acresce que, existem ainda, outros recursos de pesquisa, mais gerais, como as bases de dados científicas, entre elas destacam-se a Medical Literature Analysis and Retrieval System Online (MEDLINE), a PubMed e a Cumulative Index to Nursing and Allied Health Literature (CINAHL), que indexam um conjunto de revistas científicas, as quais publicam informação de referência e disponibilizam informação aos clínicos.

A colaboração Cochrane

A Colaboração Cochrane, que pode ser visitada em www.cochrane.org, é uma organização internacional que tem por objetivo principal ajudar as pessoas a tomar decisões bem informadas sobre os cuidados de saúde, através da preparação, manutenção e promoção da acessibilidade à evidência. Ao fornecer uma síntese confiável da evidência disponível num determinado campo do conhecimento, as RSL orientam-se sob o princípio de que a ciência é cumulativa e facilita as decisões, considerando todas as evidências sobre o efeito de uma intervenção.

Desde que foi fundada em 1993, a colaboração Cochrane tem crescido incluindo mais de 15,000 colaboradores de mais de 100 países. A colaboração internacional foi lançada um ano após a criação do Centro Cochrane em Oxford (agora o Centro Cochrane do Reino Unido), tendo sido fundada por Sir Lain Chalmers e colegas, e sendo designada mais tarde com o nome do epidemiologista britânico Archie Cochrane (Higgins & Green, 2011).

A estrutura da colaboração Cochrane organiza-se em torno de 53 grupos de revisão Cochrane (GRC), responsáveis pela preparação e manutenção de revisões dentro de áreas específicas de cuidados de saúde. Os membros desses grupos incluem investigadores,

profissionais de saúde e pessoas que utilizam os serviços de saúde (consumidores), os quais compartilham um entusiasmo comum para gerar evidência para a prevenção e tratamento de problemas ou grupos de problemas de saúde específicos.

Os GRC são suportados por grupos de metodologia, centros e campos. Os grupos de metodologia oferecem um fórum para que os metodólogos possam discutir o desenvolvimento, avaliação e aplicação de métodos utilizados para preparar as revisões Cochrane. Desempenham também um papel importante na produção e atualização do manual Cochrane de revisões sistemáticas de intervenções. Os centros Cochrane estão localizados em diferentes países e, juntos, representam todas as regiões e fornecem treino e apoio aos autores e aos grupos de revisão no desenvolvimento de revisões. Os campos Cochrane concentram-se em dimensões alargadas dos cuidados de saúde (por exemplo, cuidados primários), no tipo de consumidor (por exemplo, crianças), ou no tipo de medicamento (por exemplo, vacinas). As pessoas associadas aos campos ajudam a garantir que as prioridades e perspetivas na sua esfera de interesse são refletidas no trabalho dos GRC.

A colaboração Cochrane incentiva o envolvimento dos consumidores de cuidados de saúde, quer como parte da equipa de avaliação ou do processo editorial. O envolvimento do consumidor, cujo ponto de vista e perspetiva difere, muitas vezes, da dos profissionais de saúde e investigadores, ajuda a garantir que as revisões refletem adequadamente a variabilidade nos valores e condições das pessoas, e as circunstâncias de cuidados de saúde em diferentes países e contextos (Higgins & Green 2011).

A colaboração Cochrane disponibiliza os seguintes recursos:

- *Cochrane Database of Systematic Reviews* (CDSR) - contém textos completos (incluindo métodos, resultados e conclusões) de revisões e protocolos Cochrane;
- *Database of Abstracts of Reviews of Effects* (DARE) - contém apreciações críticas e resumos estruturados de outras revisões sistemáticas, em conformidade com critérios de qualidade explícitos, sendo organizada e mantida pelo centro de revisões e divulgação em York, Reino Unido;
- *Cochrane Central Register of Controlled Trials* (CENTRAL) - contém informações bibliográficas sobre centenas de milhares de estudos, incluindo os publicados em atas de conferências e muitas outras fontes atualmente não incluídas em outras bases de dados bibliográficas;
- *Cochrane Methodology Register* (CMR) - contém informação bibliográfica sobre artigos e livros sobre a ciência da revisão, e um registo prospetivo de estudos metodológicos.

O Instituto Joanna Briggs

O JBI é uma organização global sem fins lucrativos comprometida com o desenvolvimento, promoção e apoio à PBE. Tem como missão promover a prática de CSBE no mundo, por ser uma organização internacional líder para a tradução, transferência e utilização de evidências da aplicabilidade, adequação ao contexto, significado para o utente e eficácia das práticas de cuidados de saúde. Assim, disponibiliza os seguintes recursos:

- O JBI Connect+ é um portal em linha que fornece um conjunto de recursos para estabelecer, monitorizar e manter os CSBE (JBI, 2013a). Dentro do portal a informação está organizada num conjunto de nós ou áreas como cuidados intensivos, cuidados de idosos, cuidados a queimados, doenças crónicas, controlo da infeção, cuidados maternos, etc. Disponibiliza ainda a informação sumariada e consumível referida anteriormente.
- A JBI Database of Systematic Reviews and Implementation Reports é uma revista em linha, com arbitragem científica, que é editada pela Wolters Kluwer/Lippincott. Esta revista publica protocolos e relatórios de RSL, e relatórios de implementação de evidência na prática, que seguem as diversas metodologias do JBI. É indexada na MEDLINE, Excerpta Medica Database (Embase), Scopus, Mosby's Index (Elsevier), CINAHL, estando candidata ao Journal Citation Reports.

Programas JBI para a síntese e implementação da evidência

A JBI oferece o JBI *Comprehensive Systematic Review Training Program* (CS RTP) e o JBI *Evidence Based Clinical Fellowship Program* para preparar os clínicos para a síntese e implementação da evidência na prática. Ambos podem ser consultados em: <http://joannabriggs.org/jbi-education.html>.

Na última década, o JBI tem vindo a oferecer o programa de treino CS RTP que visa preparar os revisores para realizar revisões sistemáticas de evidências usando a abordagem JBI. O programa de cinco dias é presencial e aborda três módulos iniciais obrigatórios. Entre eles enumeram-se: (i) Introdução à PBE em saúde e RSL; (ii) Avaliação, extração e agregação de dados quantitativos a partir de estudos primários experimentais, não experimentais, de diagnóstico e prognóstico; (iii) Avaliação, extração e agregação de dados qualitativos a partir de estudos primários qualitativos e de documentos com opinião de peritos. Os restantes módulos referentes a outras metodologias de RSL podem ser realizados através de um programa de formação à distância.

O JBI oferece, também, um programa de treino para formar formadores de revisores acreditados para desenvolverem o CS RTP localmente nos respetivos centros colaboradores sediados nos vários países.

No que respeita ao treino em implementação de evidência na prática, o JBI *Evidence Based Clinical Fellowship Program* tem atualmente mais de 120 membros de vários países, entre os quais Austrália, Estados Unidos, Arábia Saudita, Reino Unido, Singapura, China, Brasil, Chile, Coreia, Taiwan, Etiópia, Quênia, Gana, Malawi, e Uganda. O JBI *Evidence Based Clinical Fellowship Program* é um dos mais bem-sucedidos programas do instituto e tem uma abordagem interdisciplinar e internacional para a preparação de um quadro de líderes especializados na promoção e apoio à adoção de abordagens baseadas em evidências na saúde.

CAPÍTULO II

SÍNTESE DA EVIDÊNCIA

Dada a explosão do conhecimento, cerca de dois milhões de novas entradas a cada ano, é difícil aos clínicos desenvolverem processos de síntese da ciência, pela perícia e recursos que um processo desta natureza exige. Por esta razão devem existir equipas especializadas e treinadas para utilizar métodos rigorosos, com vista a reunir e avaliar todas as evidências disponíveis em determinados tópicos de interesse, aliviando os clínicos deste fardo (Pearson et al., 2014). É também por esta razão, que, dentro do ciclo de translação da ciência daremos, neste texto, mais ênfase ao processo da síntese da evidência.

Há vários fatores que podem motivar os autores a proceder à síntese de evidência: clarificação de evidências conflituantes, abordagem de questões onde a prática clínica é incerta, exploração de variações na prática, confirmação da adequação das práticas correntes, ou destaque de necessidade de investigações futuras (Higgins & Green, 2011).

Uma das formas mais aceites para desenvolver síntese da ciência é a RSL, também chamada de síntese da ciência. Há, assim, um número crescente de colaborações especializadas, institutos e centros com pessoal qualificado para treinar revisores, realizar revisões sistemáticas e facilitar a cooperação entre os colaboradores (Pearson et al., 2014).

A RSL é uma investigação secundária complexa, detalhada e reproduzível que envolve um compromisso significativo de tempo e outros recursos. Nesta metodologia é utilizado um processo para localizar e sintetizar/agregar, a partir da literatura primária, todas as evidências existentes num determinado tópico, e por isso, o revisor usa uma fonte secundária de dados (Pearson et al., 2014). A RSL é uma tentativa de integrar os dados empíricos, provenientes dos estudos primários, com a finalidade de descobrir a evidência internacional e produzir declarações que devem orientar a tomada de decisões clínicas. Requer comunicação explícita e exaustiva dos métodos utilizados (Higgins & Green, 2011; JBI, 2014a).

Mesmo quando a evidência é limitada ou inexistente, as RSLs resumem as melhores evidências disponíveis sobre um tema específico, fornecendo a melhor evidência para informar a tomada de decisões e para futuras necessidades de investigação clínica (JBI, 2014a). Uma vez que o objetivo é sintetizar a melhor evidência disponível, no caso de não existir outro tipo de informação primária, a opinião de peritos pode constituir esta melhor evidência disponível (Pearson et al., 2014).

Os métodos explícitos, sistemáticos e reproduzíveis utilizados na RSL têm por objetivo minimizar diferentes tipos de viés, proporcionando, assim, resultados mais fiáveis que sustentem as conclusões e as decisões tomadas. A RSL é: uma metodologia reproduzível e explícita; uma procura sistemática que tenta identificar todos os estudos que satisfazem os critérios de elegibilidade; uma avaliação da validade das conclusões dos estudos incluídos, por exemplo, através da avaliação do risco de viés; uma apresentação sistemática, e sintetizada das características e resultados dos estudos incluídos (JBI, 2014a; Pearson et al., 2014).

Assim, para qualquer RSL devem desenvolver-se os seguintes passos: formular uma pergunta de revisão; definir critérios de inclusão e exclusão; localizar os estudos; seleccionar os estudos; avaliar a qualidade metodológica dos estudos; extrair os dados; analisar/resumir e sintetizar os resultados relevantes; apresentar os resultados; interpretar os resultados e determinar a aplicabilidade dos resultados (JBI, 2014a).

Para orientar uma revisão deverá ser previamente desenvolvido um protocolo orientador e ajustado à tipologia de revisão. Um protocolo de RSL deverá ser composto por diferentes secções como o título, informações sobre a autoria, *background*, critérios de inclusão e exclusão para a seleção dos estudos, e por métodos de pesquisa para identificação de estudos, de avaliação crítica da sua qualidade, bem como de extração e síntese dos dados. No *background* devem ser apresentados os argumentos que sustentam a necessidade de desenvolver a revisão e definidos os conceitos em análise. Devem ainda ser colocados em contexto os diversos elementos do PICO ou da respetiva adaptação específica para cada revisão. A secção de métodos que especifica os critérios de inclusão e exclusão para a seleção dos estudos inclui o tipo de estudos, tipo de participantes, tipo de intervenções, tipo de medidas de resultado, sendo que os critérios a especificar podem variar dependendo do tipo de RSL que está a ser planeada (JBI, 2014a).

Após realizar a RSL, os revisores apresentam um relatório cuja estrutura deverá ser a inerente ao desenvolvimento das secções do protocolo. A mesma deverá conter, genericamente: (i) dados respeitantes ao risco de viés dos estudos incluídos e decorrentes do processo da análise da qualidade metodológica; (ii) descrição dos estudos incluídos na revisão; (iii) resultados que dão resposta à questão de revisão; (iv) discussão dos resultados, indicação das fragilidades da revisão e dos estudos incluídos, bem como, as implicações para a prática e para a investigação; e, finalmente, (v) uma conclusão relativa à questão de revisão.

A noção de métodos para minimizar o viés e estabelecer a credibilidade nas revisões sistemáticas tem sido amplamente desenvolvida e debatida em termos de evidência quantitativa e qualitativa. É sobre esses aspetos que nos debruçaremos de seguida.

2.1 Minimizar o risco de viés no processo de síntese da evidência

Um passo crítico em qualquer RSL é a avaliação crítica da qualidade dos estudos que atendam aos critérios de inclusão da revisão em causa. Neste contexto, o método (em sentido lato) é o garante da qualidade dos resultados de investigação. Etimologicamente provém dos termos gregos *Meta*, que quer dizer além ou em direção a, e *Odos*, que significa caminho. Método pode significar instrumento, recurso, projeto, estratégia, plano, sentido de guia, orientação, modo de agir para se conseguir algo, mas também significa o processo ou conjunto de procedimentos para se chegar ao conhecimento (Apóstolo & Gameiro, 2005).

Na área da saúde o objeto de investigação é plural e, portanto, constituído por um conjunto de dimensões que se interpenetram. Este pressuposto é suportado na perspetiva pluralista popperiana. Popper (1992, 1993, 1997) considera três mundos primários de natureza distinta, embora profundamente interpenetrados, formando diversos submundos, e interativos. O Mundo 1 (físico) é o das forças físicas, dos corpos e dos estados; o Mundo 2 (psíquico) é o das emoções e dos processos psíquicos, das experiências subjetivas; o Mundo 3 é o do produto da criatividade do homem, dos artefactos, da arte, das relações interpessoais. Estes três mundos estão interligados entre si e interpenetram-se, constituindo uma realidade complexa. Deste modo, Popper (1997) contrapõe-se ao dualismo cartesiano e assume-se como pluralista, todavia, rejeitando como alternativa o monismo ontológico. Esta perspetiva, sobretudo no que respeita à interação dos mundos, abre uma multiplicidade de possibilidades para se considerar a natureza da realidade que interessa à ciência. Suporta e orienta, para além das abordagens quantitativas orientadas pelo paradigma de cariz positivista-empírico analítico, as abordagens qualitativas, como a fenomenológica, interacionista, etnográfica, entre outras orientadas pelo paradigma empírico-compreensivo.

Um paradigma pode ser considerado como uma visão ou estrutura do pensamento e consiste nas crenças básicas sobre os fenómenos que dizem respeito à disciplina. A seleção de um quadro paradigmático adequado é fundamental para os investigadores. A ontologia, a epistemologia e a metodologia moldam as dimensões de qualquer paradigma.

No pensamento ocidental são considerados três perspetivas paradigmáticas dominantes:

- 1) A perspetiva empírico-analítica, também referenciada na literatura da especialidade como positivista/racionalista, das ciências nomotéticas, ou ainda como paradigma forte. Na perspetiva positivista: previsão, controlo, verificação empírica, controlo/

domínio do mundo externo, leis universais, variáveis, e matemática são assunções fundamentais. Adequa-se ao estudo da realidade, numa perspetiva explicativa que acomete para um tipo de causalidade externa e leis universais representadas por relações constantes entre os fenómenos observados, as quais explicam o funcionamento do real. O investigador recorre à estatística que lhe permite verificar a generalidade de um fenómeno, ao mesmo tempo que pode negligenciar a variabilidade dos casos particulares (Apóstolo & Gameiro, 2005).

- 2) A perspetiva hermenêutica/compreensiva, que é também conhecida como interpretativista/naturalista ou paradigma da simultaneidade orienta para a compreensão dos fenómenos através dos significados que as pessoas lhes atribuem. As explicações para eventos sociais e culturais são baseadas nos pontos de vista e experiências das pessoas estudadas, e, considerando a natureza social do significado, existem múltiplas realidades socialmente construídas. Além disso, é fundamental orientar a investigação para a causalidade interna, que se refere ao significado dos fenómenos compreendidos num contexto específico. Contrapondo-se ao positivismo, esta posição paradigmática encaminha para a contextualização dos fenómenos humanos, onde a compreensão confronta a explicação do homem como um ser em construção (Apóstolo & Gameiro, 2005).
- 3) A teoria crítica, está historicamente relacionada com os teóricos críticos da Escola de Frankfurt que desenvolveram o primeiro esforço sistemático para empregar as técnicas de investigação empírica tradicionais para o refinamento e teste das proposições derivadas da tradição marxista (Held, 1980; Hoffman, 1989). Entre outros, Max Horkheimer, Theodor Adorno, Herbert Marcuse, Erich Fromm e mais recentemente Jürgen Habermas são reconhecidos como os mais importantes. Horkheimer um dos fundadores da Escola de Frankfurt e da teoria crítica, que assumiu a direção do Instituto de Pesquisa Social em 1930 (sendo este o berço da referida escola), considera a teoria crítica como a procura de emancipação humana para libertar o ser humano, a partir das circunstâncias que o escravizam. A teoria crítica está particularmente preocupada com a questão das relações de poder dentro da sociedade, com a interação étnica, política, económica, de classe, género, educação, religião e outras instituições sociais que contribuem para um sistema social. Em contraste com a teoria tradicional, que explora e confirma o *status quo*, a teoria crítica desafia o *status quo*, esforçando-se para o desenvolvimento de uma sociedade equilibrada e democrática e procurando a emancipação e a mudança (Held, 1980; Hoffman, 1989).

O objetivo da avaliação crítica é estimar até que ponto os riscos potenciais de viés e de consistência da qualidade foram minimizados durante a conceptualização e a realização dos estudos primários individuais, e se há utilização adequada do método e da metodologia.

Cada estudo elegível deve ser avaliado com base num conjunto de critérios para estabelecer a validade e a confiabilidade do processo e dos resultados. Esta avaliação crítica deve ser apresentada no relatório, na secção de resultados da revisão, sendo acompanhada por uma discussão relativa à qualidade metodológica e ao potencial risco de viés de todos os estudos incluídos e excluídos com base neste juízo crítico.

Quando a evidência produzida gera recomendações para a prática ou política, a inclusão de evidências de baixa qualidade e com elevado risco de viés é problemática. Nas revisões JBI a equipa de revisores deve definir um ponto de corte que oriente a decisão de excluir estudos considerados de baixa qualidade ou com alto risco de viés. Podem considerar, contudo, abordagens alternativas, tais como meta-regressão ou análise de sensibilidade, no caso de haver lugar a meta-análise (JBI, 2014a).

As evidências sobre os efeitos das intervenções podem ser constituídas a partir de estudos com risco diversificado de viés. Há duas abordagens que são igualmente razoáveis na condução de uma revisão sistemática de eficácia: a de incluir somente estudos com baixo ou moderado risco de viés, justificando como este risco é determinado, o que se considera como o grau de risco de viés e excluindo todos os estudos considerados de alto risco; e a de incluir todos os estudos, independentemente do seu risco de viés, e explicitamente considerar o risco dos diferentes estudos durante a síntese de dados (Tufanaru, Munn, Stephenson, & Aromataris, 2015).

Nas revisões Cochrane, o potencial risco de viés de cada estudo e do conjunto dos estudos é considerado como sendo baixo, alto ou pouco claro. Esta última categoria traduz falta de informação ou incerteza sobre o potencial viés do estudo primário. A avaliação é feita através do instrumento de risco de viés em relação a cada um dos seguintes processos: seleção, desempenho, deteção, atrito e publicação - *Cochrane Risk of Bias Tool* (Higgins & Green, 2011).

2.1.1 Minimizar o risco de viés numa revisão quantitativa.

A avaliação da qualidade dos estudos quantitativos de eficácia incluídos numa revisão deve enfatizar o risco de viés dos seus resultados, ou seja, o risco de que eles vão superestimar ou subestimar o efeito da intervenção (Higgins & Green, 2011). O viés não deve ser confundido com imprecisão. O viés refere-se ao erro sistemático e a precisão refere-se ao erro aleatório e reflete-se no intervalo de confiança em torno da estimativa de efeito da intervenção de cada estudo e no peso dado aos resultados de cada estudo numa meta-análise. Os resultados mais precisos têm mais peso na meta-análise (Higgins & Green, 2011; JBI 2014a).

A colaboração Cochrane estabelece a diferença entre risco de viés e qualidade metodológica. O termo avaliação da qualidade metodológica tem sido utilizado para fazer referência à avaliação crítica dos estudos incluídos e sugere que o estudo foi desenvolvido sob os mais elevados padrões metodológicos (Higgins & Green, 2011).

De facto, um estudo pode ser realizado com os mais altos padrões possíveis de qualidade, mas ter um elevado risco de viés relacionado, por exemplo, com a impraticabilidade de ocultar aos participantes a intervenção a que estão a ser sujeitos. Esta é, de facto, uma condicionante das intervenções não farmacológicas comparadas com os cuidados usuais (exemplo da estimulação cognitiva), contrariamente a intervenções farmacológicas onde é possível fazer estudos cegos com um fármaco e um placebo. Outro exemplo é o de uma intervenção cirúrgica em que é impossível ocultar essa intervenção. Assim, é inadequado julgar estes estudos como tendo baixa qualidade, não significando, contudo, que estão livres de viés resultante do conhecimento pelos participantes da intervenção a que estão a ser sujeitos (Higgins & Green, 2011).

O risco de viés nos resultados de cada estudo, que contribui para uma estimativa do efeito, é um dos vários fatores que devem ser considerados ao avaliar a qualidade de um corpo de evidência. Nos estudos de eficácia o controlo do risco de viés orienta-se para as características dos estudos em relação à seleção, intervenção, deteção, atrito, e publicação, entre outros.

Relativamente ao viés de seleção, a randomização é a única maneira de controlar as diferenças sistemáticas entre as características de base dos participantes em diferentes grupos de intervenção, em termos de fatores *confundentes* conhecidos e desconhecidos (ou não medidos). A evidência empírica sugere que, em média, os estudos não randomizados podem produzir estimativas de efeito que indicam benefícios mais extremos dos efeitos de cuidados de saúde do que os estudos randomizados.

O controlo do viés de seleção deve atender a procedimentos relativamente aos critérios de elegibilidade dos participantes, a sua randomização a partir da população e a sua alocação aos grupos (experimental ou controlo). A ocultação da alocação é um aspeto fundamental para minimizar o risco de viés. Pode levar a estimativas exageradas de efeito do tratamento, em média, mas com possibilidade de viés em qualquer direção (Higgins & Green, 2011; JBI, 2014a; Schulz & Grimes, 2002). Embora a randomização elimine o viés sistemático, não produz necessariamente grupos perfeitamente equilibrados em matéria de fatores de prognóstico. Diferenças devido ao acaso podem permanecer nos grupos de intervenção. Por essa razão, os investigadores devem apresentar as distribuições das características de base, por grupo de tratamento, numa tabela. Estas informações descrevem a população hipotética que a amostra em estudo representa (Schulz & Grimes, 2002).

Alguns programas informáticos utilizados nestes processos geram números pseudoaleatórios como é o caso do *Statistical Package for the Social Sciences*. Os programas informáticos, geralmente, fornecem números aleatórios que são gerados de acordo com um algoritmo com frequência de comportamento muito próxima, sendo que neste caso, os números não são verdadeiramente aleatórios. Tais números são chamados de pseudoaleatórios e muitas vezes gerados para o intervalo de 0,1 (de Smith, 2015). Por definição, os verdadeiros números aleatórios são gerados a partir de uma fonte de entropia e processados através de um computador sendo, assim, totalmente imprevisíveis. Por outro lado, os números pseudoaleatórios não dependem de uma fonte de entropia, mas baseiam-se num algoritmo matemático ou são gerados a partir de uma lista anteriormente calculada. Portanto, se o algoritmo ou a lista são identificados, os números gerados são predizíveis (Foley, 2001). O RANDOM.ORG (<https://www.random.org>) é um exemplo de um gerador de verdadeiros números aleatórios em que a aleatoriedade vem do ruído atmosférico.

O viés de desempenho refere-se a diferenças sistemáticas entre os grupos na exposição a outros fatores para além de intervenções, processos ou procedimentos de interesse. Após a inscrição no estudo, a ocultação aos participantes do tipo de intervenção que estão destinados a realizar pode reduzir o risco de viés (Higgins & Green, 2011; JBI 2014a). Como referimos, esta ocultação nem sempre é possível. Por exemplo, é geralmente impossível *cegar* as pessoas relativamente a determinados procedimentos como é o caso de uma cirurgia.

O viés de deteção refere-se a diferenças sistemáticas na forma como os resultados são medidos entre os diferentes grupos de tratamento. A ocultação/mascaramento aos avaliadores do tipo de intervenção que foi recebida por cada um dos grupos pode ser especialmente importante no processo de avaliação dos resultados, minimizando a subjetividade do avaliador e reduzindo o risco de enviesamento. Também a qualidade (precisão e validade) dos instrumentos ou critérios de avaliação tem um importante impacto na deteção dos efeitos da intervenção. Enquadra-se ainda neste aspeto a qualidade da análise e dos procedimentos estatísticos utilizados nos estudos primários (Higgins & Green, 2011; JBI 2014a).

O viés de atrito refere-se a diferenças sistemáticas entre os grupos relativamente às perdas de amostra desde a randomização para cada grupo à avaliação pós-intervenção ou de *follow-up*. Nestas situações os autores dos estudos primários devem reportar quais os procedimentos utilizados para minimizar o risco de viés, como proceder à análise *intention-to-treat* (ITT). Esta análise é uma estratégia que compara os grupos de estudo em termos do tratamento a que foram alocados aleatoriamente, independentemente do cumprimento integral do protocolo de tratamento. Obedece ao princípio *randomizado, tratado* ignorando os desvios de protocolo após a randomização. Apesar de algumas

críticas à tipologia de técnicas estatísticas utilizadas neste procedimento, é consensual que a eficácia clínica pode ser mal estimada se a análise ITT não for realizada (Gupta, 2011; Higgins & Green, 2011; JBI 2014a).

O viés de publicação refere-se a diferenças sistemáticas entre os resultados reportados e não reportados. Habitualmente, a decisão dos investigadores relativa à publicação dos resultados da sua investigação é feita com base na direção e significância estatística dos dados (Chan, 2008). O *funnel plot* é habitualmente utilizado para avaliar o viés de publicação. Se o viés de publicação estiver presente, os dados a partir de vários estudos deverão ser simetricamente distribuídos numa área em forma de funil. Inversamente, um gráfico de funil assimétrico indica uma amostra de estudos enviesada (Dubben & Beck-Bornholdt, 2005; Dwan, Gamble, Williamson, & Kirkham, 2013; Higgins & Green, 2011; JBI 2014a).

2.1.1.1 Instrumentos de avaliação crítica.

As diversas estruturas e colaborações que se dedicam à síntese da evidência, entre as quais a Colaboração Cochrane, a Critical Appraisal Skills Programme (CASP) e a colaboração JBI, disponibilizam uma variedade de instrumentos de avaliação dos estudos primários.

A *Cochrane Risk of Bias Tool* é utilizada nas revisões Cochrane avaliando viés de seleção, viés de desempenho, viés de deteção, viés de atrito e viés de publicação (Higgins & Green, 2011). A CASP apresenta sete instrumentos de avaliação crítica de estudos quantitativos que incluem instrumentos para revisões sistemáticas, ensaios clínicos randomizados, estudos de coorte, estudos caso-controlo, avaliações económicas, estudos de diagnóstico e normas de predição clínica. Estes instrumentos podem ser consultados em <http://www.casp-uk.net/#!checklists/cb36>.

Relativamente à avaliação da qualidade dos estudos primários de precisão de diagnóstico é de destacar a *Quality Assessment of Diagnostic Accuracy Studies* (QUADAS-2; Whiting et al., 2011). Este instrumento é recomendado para uso em revisões sistemáticas de precisão diagnóstica pela Cochrane, JBI, AHRQ e pelo U.K. National Institute for Health and Clinical Excellence.

O JBI (2015b) apresenta um instrumento para estudos de precisão diagnóstica baseado na abordagem QUADAS-2. Também a nossa descrição de cada item do instrumento para estudos de precisão diagnóstica do JBI, que apresentaremos adiante foi informada pelas considerações de Whiting et al. (2011).

Os instrumentos de avaliação crítica que apresentamos de seguida foram traduzidos, e mais tarde atualizados, considerando a informação disponibilizada em <http://joannabriggs.org/research/critical-appraisal-tools.html>. Na explicação de alguns itens foi adicionada informação para tornar mais compreensível os aspetos a avaliar. Apesar deste processo ter sido desenvolvido com o apoio de elementos do Portugal Centre for Evidence Based Practice: a JBI Centre of Excellence, as versões aqui apresentadas não devem ser consideradas como oficiais em português europeu do JBI. Este processo será desenvolvido posteriormente.

Durante o processo de avaliação crítica, em relação a cada item, de cada grelha, os revisores têm quatro opções que devem considerar: *sim*, *não*, *não claro* ou *não aplicável*.

2.1.1.1 Instrumento de Avaliação Crítica para Estudos Randomizados Controlados (ERC; JBI, 2016).

O instrumento de avaliação crítica para ERC é composto por 13 itens.

1. A alocação dos participantes aos grupos de tratamento foi verdadeiramente aleatória?

As diferenças entre os participantes incluídos nos grupos em comparação constituem uma ameaça à validade interna de um estudo que explore relações causais. Se os participantes não são alocados aos grupos de tratamento e controlo por atribuição aleatória, existe risco de que este processo seja influenciado pelas características conhecidas dos participantes e que essas diferenças possam distorcer a comparabilidade entre grupos. Uma verdadeira atribuição aleatória de participantes nos grupos significa que é utilizado um procedimento que aloca os participantes a grupos puramente baseados no acaso, não sendo influenciado pelas características conhecidas dos participantes. Verifique os detalhes sobre o procedimento de randomização utilizado para a alocação dos participantes nos grupos em estudo. Foi utilizado um procedimento verdadeiramente ao acaso (aleatório)? Por exemplo, foi utilizada uma lista de números aleatórios? Foi utilizada uma lista de números aleatórios gerada por computador?

2. A alocação aos grupos foi cega?

Se quem aloca os participantes nos grupos em comparação está consciente de qual o grupo que será o próximo no processo de alocação, isto é, tratamento ou controlo, existindo o risco de que possa intervir deliberadamente e intencionalmente na alocação

de participantes ao atribuir os mesmos de forma preferencial e, portanto, manipular a implementação do processo de alocação indicado pela randomização, podendo distorcer os resultados do estudo. A ocultação da alocação refere-se a procedimentos que impedem aqueles que alocam participantes no grupo de tratamento ou no grupo controlo de saber, antes da alocação, se o tratamento ou controlo são os próximos no processo. Verifique os detalhes relativos ao procedimento usado para a ocultação da alocação. Foi utilizado um procedimento apropriado de ocultação de alocação? Por exemplo, foi utilizada randomização central? Foram utilizados envelopes numerados sequencialmente, opacos e selados? Foram utilizados pacotes de medicamentos codificados?

3. Os grupos de tratamento eram comparáveis no início do estudo?

As diferenças entre participantes incluídos nos grupos em comparação constituem uma ameaça à validade interna de um estudo que explora relações causais. Se existirem diferenças entre os participantes incluídos nos grupos em comparação, subsiste um risco de viés de seleção. Se existirem diferenças entre os participantes incluídos nos grupos em comparação, talvez o *efeito* não possa ser atribuído à *causa* potencial, sendo plausível que o efeito possa ser explicado pelas diferenças entre participantes, ou seja, por viés de seleção. Analise as características dos participantes reportadas. Os participantes dos grupos em comparação são semelhantes em relação às características que podem explicar o efeito mesmo na ausência da causa, por exemplo, idade, gravidade da doença, estadió da doença, condições coexistentes e assim por diante? Verifique as proporções de participantes com características específicas relevantes nos grupos em comparação. Verifique os meios de medição relevantes nos grupos em comparação (score de dor, score de ansiedade, etc.). Não considere o valor-*p* para o teste estatístico de diferenças entre os grupos relativamente às características iniciais.

4. Foi ocultada aos participantes a atribuição do tratamento?

Se os participantes estão cientes da sua alocação no grupo de tratamento ou no grupo de controlo, existe o risco de se comportarem, responderem ou reagirem de forma diferente à intervenção de interesse ou à intervenção de controlo, respetivamente, e em comparação com situações em que não estão conscientes da alocação, os resultados do estudo podem ser distorcidos. De forma a minimizar este risco, *cega-se* os participantes. Cegar os participantes refere-se a procedimentos que impedem os participantes de saber qual grupo a que são alocados. Se for utilizado um procedimento para cegar

os participantes, estes não sabem se estão no grupo que recebem o tratamento de interesse ou se estão em qualquer outro grupo que recebe as intervenções de controlo. Verifique os detalhes relatados no artigo sobre o cegar dos participantes em relação à atribuição do tratamento. Foi utilizado um procedimento apropriado para este efeito? Por exemplo, foram utilizadas cápsulas ou seringas idênticas? Foram utilizados dispositivos semelhantes? Esteja consciente da utilização de diferentes termos, cegar é, também, referido como mascarar.

5. Foi ocultado aos responsáveis por aplicar o tratamento qual o grupo a que estavam alocados os participantes?

Existe o risco de quem aplica o tratamento se comportar de forma diferente com os participantes do grupo de intervenção e do grupo de controlo caso esteja ciente da alocação dos participantes pelos grupos. Isto pode influenciar a implementação dos tratamentos em comparação e os resultados do estudo podem ser distorcidos. Cegar quem aplica o tratamento é utilizado para minimizar esse risco. Cegar quem aplica o tratamento refere-se a procedimentos que impedem de saber qual o grupo a ser tratado, ou seja, aqueles que aplicam o tratamento não sabem se estão a lidar com o grupo que recebe o tratamento de interesse ou se estão a lidar com qualquer outro grupo que receba intervenções de controlo. Verifique os detalhes relatados no artigo sobre processos de ocultação a quem aplica o tratamento no que diz respeito à atribuição do tratamento. Existe alguma informação no artigo sobre quem aplica o tratamento? Quem aplica o tratamento desconhece a alocação dos participantes aos grupos em comparação?

6. Foi ocultado aos avaliadores dos resultados o grupo a que estavam alocados os participantes?

Se aqueles que avaliam os resultados estão cientes da alocação dos participantes ao grupo de tratamento ou ao grupo de controlo, existe o risco de se comportarem de forma diferente com os participantes em comparação com situações em que não estão cientes da alocação do tratamento e, portanto, existe risco de que a mensuração de resultados possa ser alterada e os resultados do estudo possam ser distorcidos. Cegar os avaliadores de resultados é utilizado para minimizar esse risco. Verifique os detalhes relatados no artigo sobre a ocultação dos avaliadores de resultados no que diz respeito à atribuição do tratamento. Existe alguma informação no artigo sobre quem avalia os resultados? Quem avalia os efeitos do tratamento nos resultados desconhece a alocação dos participantes aos grupos em comparação?

7. Os diferentes grupos do estudo foram tratados de forma idêntica, com exceção da intervenção referida?

De modo a atribuir o efeito à causa (a exposição ou intervenção de interesse), e assumindo que não existe viés de seleção, não devem existir outras diferenças entre os grupos em termos de tratamentos ou cuidados recebidos, exceto a causa manipulada. Se existirem outras exposições ou tratamentos que ocorram ao mesmo tempo com a causa (o tratamento de interesse), além da causa, então, provavelmente, o efeito não pode ser atribuído à causa potencial, na medida em que se torna plausível que o efeito possa ser explicado por outras exposições ou tratamentos que ocorrem em simultâneo com a causa. Verifique as exposições relatadas ou as intervenções recebidas pelos grupos em comparação. Existem outras exposições ou tratamentos que ocorrem em simultâneo com a causa? É plausível que o efeito possa ser explicado por outras exposições ou tratamentos que ocorrem ao mesmo tempo com a causa?

8. O *follow-up* foi completado, e se não, foi abordado o uso de estratégias para colmatar a sua ausência?

Se existirem diferenças relativamente à perda para o *follow-up* entre os grupos em comparação, essas diferenças representam uma ameaça à validade interna de um estudo que explore relações causais, uma vez que essas diferenças podem fornecer uma explicação alternativa plausível para o efeito observado mesmo aquando da ausência de uma causa. Verifique se existem diferenças em relação à perda para o *follow-up* entre os grupos em comparação. Existe uma descrição do *follow-up* incompleto (número de participantes e os motivos específicos de perda)? Se existirem diferenças entre os grupos em relação à perda para o *follow-up*, procedeu-se à análise dos padrões de perda para o *follow-up*? Se existirem diferenças entre os grupos em relação à perda para o *follow-up*, existe análise do impacto da perda para o *follow-up* nos resultados?

9. Os participantes foram analisados nos grupos aos quais foram randomizados?

Existem diferentes estratégias de análise estatística disponíveis para a análise de dados obtidos através de ensaios controlados randomizados, tais como a análise Intention-to-Treat (ITT), análise por protocolo e analisados como tratados. Na análise ITT os participantes são analisados nos grupos aos quais foram randomizados. A análise ITT compara os resultados para os participantes dos grupos iniciais criados pela alocação aleatória inicial dos participantes a esses grupos. A análise ITT não está relacionada

com a validade interna de um ensaio, trata-se de um tipo de análise estatística recomendada na declaração *Consolidated Standards of Reporting Trials* (CONSORT) sobre as melhores práticas nos relatórios de ensaios e é considerada um marcador de boa qualidade metodológica da análise dos resultados de um estudo randomizado. O ITT fornece uma estimativa estatística imparcial das diferenças nos resultados observados com a atribuição aleatória de participantes nos diferentes grupos de intervenção em comparação. Verifique se a análise ITT foi realizada. Verifique os detalhes do ITT. Esta análise foi realizada conforme recomendado na literatura metodológica internacional? Consulte o Manual de Revisores da JBI para obter referências sobre as diferentes estratégias estatísticas disponíveis para a análise de dados de ensaios clínicos randomizados e a condução adequada de ITT.

10. Os resultados foram avaliados da mesma forma para todos os grupos?

Se o resultado (o efeito) não for avaliado da mesma forma nos grupos em comparação, verifica-se uma ameaça à validade interna de um estudo que explore uma relação causal, na medida em que as diferenças nas medidas de avaliação de resultados podem ser confundidas com um efeito do tratamento (a causa). Verifique se os resultados foram avaliados da mesma maneira. Foi utilizado o mesmo instrumento ou escala? Durante o mesmo tempo de avaliação?

11. Os resultados foram medidos de forma confiável?

A falta de confiabilidade das medidas de avaliação de resultados é uma ameaça que enfraquece a validade das inferências sobre a relação estatística entre a causa e o efeito estimado num estudo que explora relações causais. A falta de confiabilidade das medidas de avaliação de resultados é uma das diferentes explicações plausíveis para erros de inferência estatística relativos à existência e à magnitude do efeito determinado pelo tratamento (causa). Verifique os detalhes sobre a confiabilidade da medida, como o número de avaliadores, a formação dos avaliadores, a confiabilidade intra-avaliador e a confiabilidade inter-avaliador dentro do estudo (não por fontes externas). Esta questão é sobre a confiabilidade da medição de resultados realizada no estudo, não se trata da validade dos instrumentos/escalas de medição utilizados. Nota: duas outras importantes ameaças que enfraquecem a validade das inferências sobre a relação estatística existente entre a causa e o efeito são o baixo poder estatístico e a violação de pressupostos de testes estatísticos. Estas ameaças não são exploradas na Pergunta 11, mas sim na Pergunta 12.

12. Foi utilizada análise estatística apropriada?

A análise estatística inadequada pode causar erros de inferência estatística em relação à existência e à magnitude do efeito determinado pelo tratamento (causa). O baixo poder estatístico e a violação dos pressupostos de testes estatísticos são duas ameaças importantes que enfraquecem a validade das inferências sobre a relação estatística entre a causa e o efeito. Verifique os seguintes aspetos: se os pressupostos de testes estatísticos foram respeitados; se foi utilizada uma análise do poder estatístico apropriado; se foram utilizados tamanhos de efeito apropriados; se foram utilizados procedimentos estatísticos adequados, dado o número e o tipo de variáveis dependentes e independentes, o número de grupos em estudo, a natureza da relação entre os grupos (grupos independentes ou dependentes) e os objetivos da análise estatística (associação entre variáveis, predição, análise de sobrevivência, etc.).

13. O desenho de estudo é apropriado ao tópico em análise, e foi evidenciado algum desvio do desenho padrão de um ERC durante as fases de desenvolvimento ou análise?

Certos desenhos de ERC, como o ERC cruzado, só devem ser realizados quando apropriado. Desenhos metodológicos alternativos podem apresentar riscos adicionais de viés, se não forem considerados no desenho e na análise.

Os ensaios cruzados só devem ser realizados com pessoas com uma condição crónica e estável, onde a intervenção produz um efeito a curto prazo (por exemplo, alívio de sintomatologia). Os ensaios cruzados devem garantir que haja um período adequado entre os tratamentos durante o qual os indivíduos não recebem tratamento.

Os ERC do tipo *cluster* randomizam grupos de indivíduos, formando *clusters*. Quando se avaliam os resultados a um nível individual em ensaios de *cluster*, emergem dificuldades relativas à unidade de análise, uma vez que os indivíduos dentro de um *cluster* estão correlacionados. Tais dificuldades devem ser tidas em consideração pelos autores do estudo ao realizar as análises e, idealmente, os autores devem relatar o coeficiente de correlação *intra-cluster*.

O ERC do tipo *stepped-wedge* pode ser apropriado quando se espera que a intervenção seja mais benéfica do que prejudicial, ou devido a considerações logísticas, práticas ou financeiras na implementação de um novo tratamento/intervenção. A análise dos dados nestes ensaios deve ser conduzida adequadamente, tendo em consideração os efeitos do tempo.

2.1.1.2 Instrumento de Avaliação Crítica para Estudos Quasi-experimentais – estudos experimentais sem alocação randomizada (JBI, 2016).

O instrumento de avaliação crítica para estudos quasi-experimentais é composto por nove itens.

1. Está claro no estudo qual é a *causa* e o *efeito* - não há confusão sobre que variável vem em primeiro lugar?

A ambiguidade respeitante à relação temporal de variáveis constitui uma ameaça à validade interna de um estudo que explora relações causais. A causa (variável independente) deve ocorrer no tempo antes do efeito explorado (variável dependente). Verifique se está claro que variável é manipulada como uma causa potencial. Verifique se está claro que variável é avaliada enquanto efeito potencial da causa. É claro que a causa foi manipulada antes da ocorrência do efeito?

2. Os participantes incluídos em qualquer comparação são semelhantes?

As diferenças entre participantes incluídos nos grupos em comparação constituem uma ameaça à validade interna de um estudo que explore relações causais. Se existirem diferenças entre os participantes incluídos nos grupos em comparação, existe o risco de viés de seleção. Se existirem diferenças entre os participantes incluídos nos grupos em comparação, talvez o efeito não possa ser atribuído à causa potencial, sendo talvez plausível que o efeito possa ser explicado pelas diferenças entre participantes, ou seja, por viés de seleção. Verifique as características reportadas para os participantes. Os participantes dos grupos em comparação são semelhantes em relação às características que podem explicar o efeito mesmo na ausência da causa, por exemplo, idade, severidade da doença, estadió da doença, condições coexistentes, entre outros? Nos estudos pré-teste/pós-teste, onde os utentes são os mesmos em qualquer comparação, a resposta a esta pergunta deverá ser sim.

3. Os participantes incluídos receberam tratamento/cuidados semelhantes, além da exposição ou intervenção de interesse?

De modo a atribuir o efeito à causa (a exposição ou intervenção de interesse), assumindo que não existe viés de seleção, não devem existir diferenças entre os grupos a nível de

tratamentos ou cuidados recebidos, exceto a causa manipulada. Se existirem outras exposições ou tratamentos que ocorram no mesmo período que a causa, além desta, então o efeito potencial não pode ser atribuído à causa potencial, pois é plausível que o efeito possa ser explicado por outras exposições ou tratamentos que ocorram ao mesmo tempo que a causa. Verifique as exposições ou as intervenções reportadas para os grupos em comparação. Existem outras exposições ou tratamentos que ocorram ao mesmo tempo que a causa? É plausível que o efeito possa ser explicado por outras exposições ou tratamentos que ocorram ao mesmo tempo que a causa?

4. Existe um grupo de controlo?

Os grupos de controlo oferecem as condições necessárias para explorar o que teria acontecido com grupos expostos a outros tratamentos diferentes que não a potencial causa. A comparação do grupo de tratamento (o grupo exposto à causa em análise) com outros grupos reforça a análise da probabilidade causal. A validade das inferências causais é reforçada em estudos com um grupo controlo independente comparativamente com estudos sem grupo de controlo independente. Este grupo de controlo deverá ser separado e independente (não o pré-teste num estudo com desenho de um único grupo).

5. Existiram múltiplas avaliações de resultado antes e após a intervenção/exposição?

De modo a evidenciar a existência de mudanças no resultado (o efeito) como consequência da intervenção/tratamento (a causa), é necessário comparar os resultados da avaliação antes e depois da intervenção/tratamento. Se não existir avaliação antes do tratamento e apenas estiver disponível a avaliação após o tratamento, não se sabe se há alteração após o tratamento em relação ao antes do tratamento. Se forem recolhidas múltiplas avaliações antes da implementação da intervenção/tratamento, torna-se possível explorar a probabilidade de explicações alternativas que não a causa proposta para o efeito observado, tais como as alterações de ocorrência natural na ausência da causa e mudanças de pontuação alta (ou baixa) para valores menos extremos, mesmo na ausência da causa (às vezes referida como regressão para a média). Se forem realizadas várias avaliações após a implementação da intervenção/tratamento, é possível explorar as alterações do efeito ao longo do tempo em cada grupo e compará-las entre os grupos. Verifique se foram realizadas avaliações antes da implementação da intervenção/tratamento. Verifique se foram realizadas avaliações após a implementação da intervenção/tratamento. Existem múltiplas medidas de avaliação no pré-teste? Existem múltiplas medidas de avaliação no pós-teste?

6. O período de *follow-up* foi completo e, se não, o *follow-up* foi adequadamente descrito e utilizadas estratégias para lidar com a perda para o *follow-up*?

Se existirem diferenças quanto à perda de acompanhamento entre os grupos em comparação, estas representam uma ameaça à validade interna de um estudo que explore efeitos causais, na medida em que estas diferenças podem providenciar uma explicação alternativa plausível para o efeito observado na ausência da causa. Verifique se existem diferenças em relação à perda para o *follow-up* entre os grupos em comparação. Existe uma descrição relativa ao *follow-up* incompleto (número de participantes e as razões específicas para a perda para o *follow-up*)? Se existirem diferenças entre os grupos em relação à perda para o *follow-up*, foi incluída a análise dos padrões de perda para o *follow-up*? Se existirem diferenças entre os grupos no que se refere à perda para o *follow-up*, é analisado o impacto desta perda nos resultados?

7. Os resultados dos participantes incluídos foram avaliados de igual forma?

Se o resultado (o efeito) não é avaliado da mesma forma nos grupos em comparação evidencia-se uma ameaça à validade interna de um estudo que explore relações causais, na medida em que as diferenças nas avaliações dos resultados podem ser confundidas com um efeito do tratamento (a causa). Verifique se os resultados foram avaliados da mesma forma. Foi utilizado o mesmo instrumento ou escala? No mesmo momento de avaliação?

8. Os resultados foram avaliados de forma confiável?

A falta de confiabilidade das medidas de avaliação de resultados é uma ameaça que enfraquece a validade das inferências sobre a relação estatística entre a causa e o efeito estimado num estudo que explore efeitos causais. A falta de confiabilidade das medidas de avaliação de resultados é uma das diferentes explicações plausíveis para erros de inferência estatística em relação à existência, e magnitude, do efeito determinado pelo tratamento (causa). Verifique os detalhes sobre a confiabilidade da medida de avaliação, tais como o número de avaliadores, a formação dos avaliadores, a confiabilidade intra-avaliador e a confiabilidade inter-avaliador dentro do estudo (não para fontes externas). Esta questão prende-se com a confiabilidade da avaliação realizada no estudo, não se trata da validade dos instrumentos/escalas de avaliação utilizados no estudo. Nota: Duas outras importantes ameaças que enfraquecem a validade das inferências sobre a relação estatística entre a causa e o efeito são o baixo poder estatístico e a violação dos pressupostos de testes estatísticos. Estas ameaças não são exploradas na Pergunta 8, mas sim na Pergunta 9.

9. Foi utilizada análise estatística apropriada?

A análise estatística inadequada pode originar erros de inferência estatística relativos à existência, e magnitude, do efeito determinado pelo tratamento (causa). O baixo poder estatístico e a violação dos pressupostos de testes estatísticos são duas importantes ameaças que enfraquecem a validade das inferências sobre a relação estatística entre a causa e o efeito. Verifique os seguintes aspetos: se os pressupostos de testes estatísticos foram respeitados; se foi realizada análise de poder estatístico apropriada; se foram utilizados tamanhos de efeito (*effect sizes*) apropriados; se foram utilizados procedimentos estatísticos adequados, dado o número e o tipo de variáveis dependentes e independentes, o número de grupos em estudo, a natureza da relação entre os grupos (grupos independentes ou dependentes) e os objetivos da análise estatística (associação entre variáveis, predição, análise de sobrevivência, etc.).

2.1.1.1.3 Instrumento de Avaliação Crítica de Estudos de Coorte (JBI, 2016).

O instrumento de avaliação crítica é composto por 11 itens.

1. Os dois grupos foram semelhantes e recrutados da mesma população?

Verifique cuidadosamente no estudo as descrições dos participantes de modo a determinar se os indivíduos dentro e entre os grupos apresentam características semelhantes em relação à exposição. Os dois grupos selecionados para comparação devem ser tão semelhantes quanto possível em todas as características, com exceção do seu estado de exposição, relevantes para o estudo em questão. Os autores devem fornecer critérios claros de inclusão e exclusão desenvolvidos antes do recrutamento dos participantes para o estudo.

2. As exposições foram medidas de forma semelhante de modo a alocar os participantes nos grupos expostos e não expostos?

Um estudo de alta qualidade com desenho coorte deve mencionar ou descrever como foram medidas as exposições em estudo. As medidas de avaliação de exposição devem ser claramente definidas e descritas em pormenor. Tal processo permitirá que os revisores avaliem se os participantes receberam, ou não, a exposição de interesse.

3. A exposição foi medida de forma válida e confiável para alocação aos grupos?

O estudo deve descrever claramente o método de medição da exposição. Avaliar a validade requer a existência de um *padrão de ouro*, com o qual a medida possa ser comparada. A validade da medição da exposição relaciona-se geralmente com a adequação da medida atual ou se uma medida de exposição anterior é necessária. Confiabilidade refere-se aos processos incluídos num estudo epidemiológico de modo a verificar a repetibilidade das medições das exposições. Estes incluem usualmente a confiabilidade intra-observador e confiabilidade entre observadores.

4. Foram identificados fatores confundentes?

O confundimento ocorre quando o efeito estimado da exposição à intervenção é influenciado pela presença de alguma diferença entre os grupos em comparação (para além da exposição investigada/de interesse). Os fatores confundentes típicos incluem características iniciais (*baseline*), fatores de prognóstico ou exposições concomitantes (exemplo: fumar). Um confundente é uma diferença entre os grupos em comparação que influencia a direção dos resultados do estudo. Um estudo coorte de alta qualidade irá identificar os potenciais fatores de confundimento e avaliá-los (sempre que possível). Esta análise torna-se difícil em estudos onde fatores comportamentais, atitudinais ou relativos ao estilo de vida dos participantes podem influenciar os resultados.

5. Foram declaradas estratégias para lidar com fatores confundentes?

Estratégias para lidar com os efeitos de fatores confundentes podem ser aplicadas no momento do desenho do estudo ou na análise de dados. Ao combinar ou estratificar a amostragem dos participantes, os efeitos dos fatores confundentes podem ser ajustados. Ao lidar com ajustes na análise de dados, avalie as estatísticas utilizadas no estudo. A maioria apresentará um tipo de análise de regressão multivariada para explicar os fatores confundentes avaliados. Esteja atento à descrição de métodos estatísticos, como métodos de regressão (sendo exemplo a regressão logística), usualmente utilizados para lidar com fatores confundentes associados a variáveis de interesse.

6. Os grupos/participantes não apresentavam o resultado de interesse no início do estudo (ou durante o momento da exposição)?

Os participantes não apresentavam o resultado de interesse no início do estudo. Consulte a secção de métodos no estudo para obter esta informação, a qual é normalmente

encontrada nas descrições do recrutamento de participantes/amostras, definições de variáveis e/ou critérios de inclusão/exclusão.

7. Os resultados foram medidos de forma válida e confiável?

Analise a secção de métodos do estudo. Se, por exemplo, a avaliação do cancro do pulmão for baseada em definições ou critérios de diagnóstico já existentes, então a resposta a esta questão é muito provavelmente sim. Se o cancro do pulmão for avaliado através de relatos observacionais, ou escalas de autopreenchimento, o risco de relato de informação em défice ou excesso aumenta, e a objetividade é comprometida. Com igual importância, determine se as ferramentas de avaliação utilizadas são instrumentos validados, dado o impacto significativo que apresentam na validade do processo de avaliação de resultados. Após ter estabelecido a objetividade da medição de avaliação (exemplo, cancro do pulmão), é fundamental descrever como é que o processo de avaliação foi desenvolvido. Os elementos envolvidos na colheita de dados foram treinados ou formados no uso do(s) instrumento(s)? (exemplo, radiologistas). Se existir mais do que um elemento na colheita de dados, eram estes semelhantes em termos do nível de educação, experiência clínica, experiência na investigação, ou nível de responsabilidade perante o trabalho de investigação?

8. O período de *follow-up* foi relatado e suficientemente longo para que os resultados ocorram?

O período de tempo adequado para *follow-up* variará de acordo com a natureza e características da população de interesse e/ou da intervenção, doença ou exposição. Para estimar uma duração apropriada, consulte diversos artigos e registe os intervalos de tempo de *follow-up*. As opiniões de peritos em prática clínica ou investigação clínica podem também ajudar a determinar uma duração adequada de acompanhamento. Por exemplo, um período temporal mais longo pode ser necessário para analisar a associação entre a exposição ocupacional ao amianto e o risco de cancro do pulmão. É importante, particularmente em estudos de coorte, que o acompanhamento seja longo o suficiente para permitir os resultados. No entanto, a questão da investigação e os resultados que estão a ser avaliados ditarão, provavelmente, o tempo de *follow-up*.

9. O *follow-up* foi completo e, se não, as razões para eventuais perdas para o *follow-up* foram descritas e exploradas?

É importante num estudo de coorte que uma percentagem significativa de participantes seja seguida. Enquanto orientação geral, pelo menos 80% dos participantes

devem ser seguidos. Geralmente, uma taxa de abandono igual ou inferior a 5% é considerada insignificante. Considera-se que uma taxa igual ou superior a 20% afeta significativamente a validade do estudo. No entanto, em estudos observacionais conduzidos durante um longo período de tempo é de esperar uma taxa de abandono superior. A decisão de inclusão ou exclusão de determinado estudo nestas condições deverá basear-se nas razões que motivaram o abandono de participantes, e se as taxas de abandono eram comparáveis entre o grupo exposto e não exposto. O relato de esforços desenvolvidos para acompanhar os participantes que abandonaram o estudo pode ser considerado um indicador de um processo bem conduzido. Procure uma descrição clara e justificável do motivo pelo qual as pessoas foram deixadas de fora, excluídas, abandonadas, etc.

10. Foram utilizadas estratégias para o *follow-up* incompleto?

Algumas pessoas podem desistir por diversos motivos como mudança de emprego ou morte. Todavia, é importante que os seus resultados sejam analisados. O viés de seleção pode ocorrer em resultado de um *follow-up* incompleto. Neste sentido, participantes com diferentes períodos de *follow-up* devem ser tidos em consideração durante a análise, a qual deverá ser ajustada de modo a permitir as diferenças temporais nos períodos de *follow-up*. Este processo é usualmente realizado através do cálculo de taxas que contabilizem o risco pessoa-anos (exemplo: considerando o tempo enquanto denominador).

11. Foi utilizada análise estatística apropriada?

Como qualquer consideração de análise estatística, deve-se ponderar se outro método estatístico alternativo mais apropriado poderia ter sido utilizado. As secções de método dos estudos de coorte devem ser detalhadas para que os revisores identifiquem quais as técnicas analíticas utilizadas (em particular, regressão ou estratificação), e como é que os fatores confundentes foram avaliados. Em estudos que utilizem a análise de regressão, torna-se útil a identificação no estudo de que variáveis foram incluídas e de que modo estas estão relacionadas com os resultados. Se a estratificação foi a abordagem analítica utilizada, os estratos de dados foram definidos de acordo com as variáveis específicas? Adicionalmente, é igualmente importante analisar a adequação das estratégias analíticas relativamente às suposições associadas à abordagem, na medida em que diferentes métodos de análise baseiam-se em suposições divergentes sobre os dados e sobre o modo como estes irão responder.

2.1.1.4 Instrumento de Avaliação Crítica para Estudos Caso-Controllo (JBI, 2016).

O instrumento de avaliação crítica para estudos caso-controllo é composto por 10 itens. Os estudos de caso-controllo comparam os resultados de grupos que receberam e não receberam uma intervenção ou uma exposição. No entanto, o método de alocação do grupo nos estudos de caso-controllo não é aleatório (JBI, 2016).

1. Os grupos eram comparáveis além da presença de doença nos casos ou ausência de doença no controlo?

O grupo de controlo deve ser representativo da população de origem que produziu os casos. Isto é habitualmente conseguido por correspondência individual, na medida em que os controlos são selecionados para cada caso com base na semelhança em certas características que não a exposição ao fenómeno em foco. A frequência ou correspondência de grupos são métodos alternativos. Se os grupos não forem comparáveis entre si pode haver viés de seleção.

2. Os casos e controlos foram agrupados apropriadamente?

Como no Item 1, o estudo deve incluir definições claras da população de origem. A origem do recrutamento dos casos e controlos deve ser cuidadosamente analisada. Por exemplo, centros de registo oncológico podem ser utilizados no recrutamento de participantes para um estudo que investiga fatores de risco no cancro do pulmão, o que tipifica os estudos de caso-controllo baseados na população. Os participantes do estudo podem ser selecionados de uma população-alvo, população de origem ou de um conjunto de participantes elegíveis (como em estudos caso-controllo a nível hospitalar).

3. Foram utilizados os mesmos critérios na identificação de casos e controlos?

Torna-se vantajoso determinar se os participantes foram incluídos no estudo baseados num diagnóstico ou definição em específico. É espetável que este método diminua o risco de viés. As características são outra abordagem vantajosa na correspondência de grupos, e estudos que não utilizem diagnósticos ou definições específicos devem apresentar evidência de correspondência pelas características chave. Um caso deve ser claramente definido. É igualmente importante que os controlos cumpram todos os critérios de elegibilidade definidos para os casos, exceto os relacionados com o diagnóstico da doença.

4. A exposição foi avaliada de forma padronizada, válida e confiável?

O estudo deve claramente descrever o método utilizado na avaliação da exposição. Aferir a validade requer que um padrão de ouro esteja disponível, face ao qual a medida de avaliação possa ser comparada. A validade da mensuração da exposição está geralmente relacionada com o facto de uma medida atual ser apropriada ou se uma medida de exposição anterior é necessária. Estudos de caso-controlo investigam diferentes tipos de exposições que podem, ou não, estar associados com a condição. Nestes casos, os revisores devem utilizar a principal exposição de interesse para a sua revisão de modo a responder a esta questão. A confiabilidade refere-se aos processos incluídos num estudo epidemiológico de modo a analisar a repetibilidade de avaliação de exposições. Habitualmente, estes processos incluem níveis de confiabilidade inter e intra-observacionais.

5. A exposição foi avaliada da mesma forma para os casos e controlos?

Como no Item 4, o estudo deve descrever claramente os métodos de mensuração da exposição. As medidas de avaliação da exposição devem ser claramente definidas e descritas em detalhe. A avaliação da exposição, ou de fatores de risco, deve ser desenvolvida de acordo com os mesmos procedimentos ou protocolos para ambos (casos e controlos).

6. Foram identificados fatores confundentes?

Ocorreu confundimento quando o efeito estimado de exposição à intervenção é tendencioso pela presença de alguma diferença entre os grupos em comparação (para além da exposição investigada/de interesse). Fatores confundentes típicos incluem características iniciais (*baseline*), fatores de prognóstico ou exposições concomitantes (ex.: fumar). Um fator confundente é a diferença entre os grupos em comparação, que irá influenciar a direção dos resultados do estudo. Um estudo de alto nível, com desenho de estudo de caso-controlo, deverá identificar os potenciais fatores de confundimento e mensurá-los (quando possível). Tal torna-se difícil para estudos onde comportamentos, atitudes ou estilos de vida podem ter impacto nos resultados.

7. Foram expressas as estratégias utilizadas para analisar fatores confundentes?

Estratégias para lidar com os efeitos dos fatores confundentes podem ser trabalhadas a nível do desenho do estudo ou na avaliação dos dados obtidos. Ao corresponder ou estratificar a amostra de participantes, os efeitos dos fatores confundentes podem ser ajustados. Ao

lidar com o ajustamento na análise de dados, avalie a estatística utilizada no estudo. A maioria será representada sob a forma de análise de regressão multivariada, de modo a ter em conta os fatores confundentes medidos. Tenha em especial atenção a descrição dos métodos estatísticos, na medida em que métodos de regressão (como a regressão logística) são usualmente utilizados para lidar com os fatores confundentes das variáveis de interesse.

8. Os resultados foram avaliados de forma padronizada, válida e credível, para os casos e controlos?

Leia a secção referente ao método no estudo. Se, por exemplo, a avaliação do cancro do pulmão for baseada em definições ou critérios de diagnóstico existentes, então a resposta a esta pergunta é provavelmente sim. Se o cancro do pulmão é avaliado através de relato observacional ou escalas autorrelatadas, aumenta o risco de notificação insuficiente ou em excesso o que compromete a objetividade. De igual importância, determine se as ferramentas de avaliação utilizadas são instrumentos validados, o que apresenta um impacto significativo na validade das avaliações de resultados realizadas.

Tendo estabelecido a objetividade do instrumento de avaliação de resultados (por exemplo, cancro de pulmão), é importante estabelecer como foi realizada a avaliação. Os elementos envolvidos na colheita de dados foram formados ou treinados na utilização do instrumento(s)? (por exemplo, radiologistas). Se existir mais do que um avaliador, estes são idênticos em termos do nível de escolaridade, experiência no âmbito clínico ou investigacional, ou nível de responsabilidade relativo à investigação a ser realizada?

9. O período de exposição foi suficientemente longo para ser significativo?

É particularmente importante num estudo caso-controlo que o tempo de exposição tenha sido suficiente para evidenciar uma associação entre a exposição e o resultado. Poderá verificar-se que o período de exposição seja demasiado curto ou prolongado, influenciando o resultado.

10. Foi utilizada análise estatística apropriada?

Como qualquer consideração de análise estatística, deve-se ponderar se outro método estatístico alternativo mais apropriado poderia ter sido utilizado. Adicionalmente, é também importante avaliar a adequação da estratégia analítica em termos dos pressupostos associados à abordagem, uma vez que os diferentes métodos de análise são baseados em distintos pressupostos sobre os dados e como irão estes responder.

2.1.1.1.5 Instrumento de Avaliação Crítica de Estudos de Caso (JBI, 2016).

O instrumento de avaliação crítica de estudos de caso é composto por oito itens.

1. As características demográficas do participante são claramente descritas?

O estudo de caso descreve claramente detalhes do participante como idade, género, raça, história clínica, diagnóstico, prognóstico, tratamentos anteriores, resultados de testes diagnósticos passados e atuais, assim como qualquer medicação? O local e contexto também podem ser descritos.

2. A história do participante foi claramente descrita e apresentada sequencialmente?

Um bom estudo de caso deverá descrever detalhadamente a história familiar, clínica e psicossocial do participante, incluindo informação genética relevante, assim como intervenções passadas indispensáveis e os seus resultados (CARE Checklist 2013; Gagnier et al., 2013).

3. A situação clínica atual do utente foi claramente descrita na apresentação?

A condição clínica atual do participante deve ser descrita em detalhe, incluindo a singularidade da condição/doença, sintomas, frequência e severidade. O relato do caso também deve indicar se foi considerado um diagnóstico diferencial.

4. Os testes ou métodos de diagnóstico, e respetivos resultados, foram claramente descritos?

Devem ser fornecidas ao leitor de um estudo de caso informações suficientes para compreender a avaliação do participante. Para tal, é importante que todos os testes apropriados sejam prescritos de modo a confirmar um diagnóstico e, portanto, o estudo deve fornecer uma descrição clara dos vários testes utilizados (padrão de ouro ou testes diagnósticos alternativos). Fotografias ou ilustrações de procedimentos diagnóstico, radiografias ou procedimentos de tratamento são usualmente apresentados quando apropriado e de modo a transmitir uma mensagem clara aos leitores.

5. A(s) intervenção(ões) ou procedimento(s) de tratamento foram claramente descritos?

É importante descrever claramente os procedimentos de tratamento ou intervenção, na medida em que outros profissionais de saúde irão analisar o documento, permitindo-lhes uma compreensão clara do protocolo de tratamento seguido. O relatório deve descrever o protocolo de tratamento/intervenção em detalhe; por exemplo, na gestão farmacológica da ansiedade dentária - o tipo de fármaco, via de administração, dosagem e frequência do fármaco, assim como quaisquer efeitos secundários.

6. A condição clínica pós-intervenção foi claramente descrita?

Um bom estudo de caso deve claramente descrever a condição clínica pós-intervenção em termos da presença ou ausência dos sintomas. Os resultados da gestão/tratamento quando apresentados como imagens ou figuras ajudam a transmitir a informação ao leitor/profissional de saúde.

7. Foram identificados e descritos eventos adversos (danos) ou eventos imprevistos?

São riscos inerentes a qualquer tratamento/intervenção/terapêutica, podendo em alguns casos constituir complicações graves. É importante que os eventos adversos sejam claramente documentados e descritos, particularmente quando está a ser tratada uma condição nova/única ou quando um novo medicamento ou tratamento é testado. Além disso, a ocorrência de eventos imprevistos, se houver, deve ser claramente descrita de modo a que possam ser extraídas informações novas ou úteis.

8. O estudo de caso fornece informações aplicáveis?

Os estudos de casos devem resumir as principais lições aprendidas em termos do conhecimento prévio da condição/doença e orientações para a prática clínica que guiam os profissionais em casos semelhantes.

2.1.1.1.6 Instrumento de Avaliação Crítica de Série de Casos (JBI, 2016).

Define-se Série de Casos como estudos onde apenas fazem parte da amostra os participantes com uma determinada doença ou resultado/desfecho. O instrumento de avaliação crítica é composto por 10 itens.

1. Existem critérios claros para inclusão na série de casos?

Os autores devem fornecer critérios de inclusão claros (e critérios de exclusão quando apropriado) para os participantes do estudo. Os critérios de inclusão/exclusão devem ser especificados (por exemplo: risco, fase de progressão da doença) com detalhes suficientes e com toda a informação essencial para o estudo.

2. A condição foi medida de forma padronizada e confiável para todos os participantes incluídos na série de casos?

O estudo deve descrever claramente o método de avaliação da condição. Esta avaliação deve ser realizada de forma padronizada (isto é, da mesma forma para todos os participantes) e fiável (isto é, resultados repetíveis e reproduzíveis).

3. Foram utilizados métodos válidos para a identificação da condição para todos os participantes incluídos na série de casos?

Diversas condições de saúde não são facilmente diagnosticadas ou definidas e alguns métodos podem não ser capazes de incluir ou excluir determinados níveis ou estadios clínicos específicos. Se os resultados foram avaliados com base em definições existentes ou critérios de diagnóstico, então a resposta a esta pergunta é provável que seja sim. Se os resultados foram avaliados com a utilização de escalas observacionais, ou escalas autorrelatadas, o risco de excesso ou sub-relato é aumentado, e a objetividade é comprometida. É igualmente importante determinar se os instrumentos de avaliação utilizados se encontram validados, dado o seu impacto significativo na legitimidade da avaliação dos resultados.

4. A inclusão de participantes na série de casos foi consecutiva?

Estudos que indiquem a inclusão consecutiva de participantes são mais confiáveis. Por exemplo, uma série de casos onde se afirme “incluímos todos os utentes (24) com osteossarcoma que se apresentaram na nossa clínica entre março de 2005 e junho de 2006” é de maior confiança do que um estudo que simplesmente afirme “relatamos uma série de casos de 24 pessoas com osteossarcoma”.

5. A série de casos teve a inclusão total dos participantes?

A integridade de uma série de casos contribui para a sua fiabilidade. Estudos que indicam uma inclusão total são mais confiáveis do que aqueles que não. Como referenciado

anteriormente, uma série de casos que afirme “incluímos todos os utentes (24) com osteossarcoma que se apresentaram na nossa clínica entre março de 2005 e junho de 2006” é de maior confiança do que um estudo que simplesmente afirme “relatamos uma série de casos de 24 pessoas com osteossarcoma”.

6. Existe descrição clara das características demográficas dos participantes no estudo?

Um estudo do tipo série de casos deve descrever claramente dados demográficos relevantes dos participantes, tais como a idade, género, nível de educação, região geográfica, etnia, período temporal, entre outros.

7. Existe descrição clara das informações clínicas dos participantes?

Informações clínicas relevantes dos participantes devem estar claramente reportadas, tais como o tipo de doença, comorbilidades, estadió da doença, intervenções/tratamentos prévios, resultados de testes de diagnóstico, etc.

8. Os resultados ou efeitos do *follow-up* dos casos foram claramente descritos?

Os resultados de qualquer intervenção ou tratamento devem ser claramente reportados na série de casos. Um bom estudo de caso deve descrever claramente a condição clínica pós-intervenção em termos da presença ou ausência de sintomas. Os resultados da gestão/tratamento, quando apresentados por imagens ou figuras, podem assistir na transmissão de informação ao leitor/profissional de saúde. É importante que eventos adversos sejam claramente documentados e descritos, particularmente aquando do tratamento de uma condição recente/única ou quando um novo medicamento ou tratamento é usado. Além disto, eventos imprevistos, se houver, que possam produzir informações novas ou úteis, devem ser identificados e claramente descritos.

9. Existe uma descrição clara da informação relativa aos contextos clínicos/demográficos apresentados?

Certas doenças ou condições variam em prevalência entre diferentes regiões geográficas e populações (por exemplo: mulheres vs. homens, variáveis sociodemográficas entre países). A amostra do estudo deve ser descrita em detalhe para que outros investigadores possam determinar se é comparável à população do seu interesse.

10. A análise estatística utilizada é adequada?

Como qualquer consideração de análise estatística, deve-se ponderar se outro método estatístico alternativo mais apropriado poderia ter sido utilizado. A secção de metodologia dos estudos deve ser suficientemente detalhada para que os revisores identifiquem quais as técnicas analíticas utilizadas e a sua adequabilidade.

2.1.1.1.7 Instrumento de Avaliação Crítica de Estudos de Prevalência (Munn, Moola, Lisy, Riitano, & Tufanaru, 2015).

O instrumento de avaliação crítica de estudos de prevalência é composto por nove itens.

1. A estrutura da amostra é representativa da população-alvo?

Esta questão baseia-se no conhecimento das características mais alargadas da população de interesse e área geográfica. Se o estudo se foca em mulheres com cancro de mama, o conhecimento de, pelo menos, características, dados demográficos e historial médico é necessário. O termo *população-alvo* não deve ser levado a inferir cada indivíduo, em qualquer lugar ou com características semelhantes de doença ou exposição. Contrariamente, considere as características específicas da população em estudo, incluindo a faixa etária, género, comorbilidades, terapêutica e outros fatores potencialmente influenciadores. Por exemplo, determinada estrutura da amostra pode não ser apropriada para representar a população-alvo se um certo grupo tiver sido utilizado (como aqueles que trabalham para uma organização ou profissão), e os resultados inferidos para a população-alvo (exemplo: adultos que trabalham). A estrutura da amostra pode ser apropriada quando inclui quase todos os membros da população alvo (exemplo: um censo, uma lista completa de participantes ou dados de registo completo).

2. Os participantes do estudo foram recrutados de forma adequada?

Os estudos podem reportar a amostragem aleatória de uma população e a secção de métodos deve relatar como este processo foi realizado. A amostragem probabilística aleatória de um subconjunto populacional definido (estrutura da amostra) deve ser empregue na maioria dos casos, contudo, não é necessária a realização de amostragem probabilística aleatória se todos os participantes na estrutura de amostragem forem incluídos/analísados. Por exemplo, reportar todos os dados de um censo é apropriado, na medida em que um bom censo irá identificar todos os indivíduos. Ao utilizar

amostragem em *cluster*, como uma amostra aleatória de aldeias dentro de uma região, os métodos precisam de estar claramente definidos, sendo que a precisão da estimativa final de prevalência incorpora o efeito de cluster. Amostras por conveniência, como um questionário de rua ou entrevistas a várias pessoas em reuniões públicas, não são consideradas como fornecedoras de uma amostra representativa da população base.

3. O tamanho da amostra é adequado?

Quanto maior a amostra, mais estreito será o intervalo de confiança em torno da estimativa de prevalência, tornando os resultados mais precisos. Um tamanho de amostra adequado é importante para garantir uma boa precisão da estimativa final. Idealmente, procuram-se evidências de que os autores realizaram o cálculo do tamanho de amostra para determinar qual o tamanho adequado. Este processo irá estimar qual o número de indivíduos necessários para produzir uma estimativa confiável da(s) medida(s) de interesse. Para condições com baixa prevalência, torna-se necessário um tamanho de amostra maior. Considere os tamanhos de amostra para a análise de subgrupos (ou características), e se estes são apropriados. Por vezes, o estudo será suficientemente grande (como estudos de investigação em âmbito nacional), pelo que não se torna necessário o cálculo do tamanho de amostra. Nesses casos, o tamanho da amostra pode ser considerado adequado.

Quando não existe cálculo do tamanho da amostra e não se está perante uma investigação de âmbito nacional, os revisores podem considerar a realização da análise do tamanho da amostra utilizando para o efeito a seguinte fórmula (Daniel 1999; Naing, Winn, & Rusli, 2006):

$$N = [Z^2 P(1-P)]/d^2$$

Onde:

n = tamanho da amostra

Z = estatística Z para um nível de confiança

P = Previsão de prevalência ou proporção (na proporção de um, se 20%, P = 0,2)

d = precisão (em proporção de um, se 5%, d = 0,05)

4. Estão os sujeitos do estudo, bem como o seu contexto, descritos detalhadamente?

Algumas doenças ou condições variam na prevalência em diferentes regiões e populações (por exemplo, mulheres vs. homens, variáveis sociodemográficas entre países). A amostra

do estudo deverá ser descrita de forma suficientemente detalhada de modo a que outros investigadores possam determinar se é comparável à sua população de interesse.

5. A análise dos dados foi realizada com cobertura suficiente da amostra identificada?

O *coverage bias* (viés de cobertura/abrangência) poderá ocorrer quando nem todos os subgrupos da amostra identificada respondem ao mesmo ritmo. Por exemplo, poderá obter uma taxa geral de resposta elevada no seu estudo, mas a taxa de resposta para um determinado subgrupo (por exemplo, adultos mais velhos) pode ser bastante baixa.

6. Os métodos utilizados para a identificação da condição são válidos?

Neste ponto, procura-se viés de medida ou de classificação. Muitas das condições de saúde não são facilmente diagnosticadas ou definidas, e algumas medidas podem não ser capazes de incluir ou excluir níveis ou estádios adequados do problema de saúde. Se os resultados foram avaliados com base em definições existentes ou critérios de diagnóstico, a resposta a esta pergunta provavelmente será sim. Se os resultados foram avaliados utilizando relatos observacionais, ou escalas autorrelatadas, o risco de excesso ou sub-relato é aumentado e a objetividade está comprometida. Igualmente importante, determine se as ferramentas de avaliação utilizadas são instrumentos validados, dado o seu impacto significativo na validade da avaliação de resultados.

7. A condição foi medida de forma confiável?

É necessário um julgamento considerável para determinar a presença de alguns resultados em saúde. Tendo estabelecido a validade do instrumento de avaliação de resultados (ver o Item 6 desta escala), é importante estabelecer como foi realizada a avaliação. Os elementos envolvidos na colheita de dados foram treinados ou formados na utilização do(s) instrumento(s)? Se existir mais do que um elemento, estes são semelhantes em termos do nível de educação, experiência clínica, investigacional ou nível de responsabilidade na investigação? Quando existia mais do que um observador ou avaliador, houve comparação de resultados entre os observadores? A condição foi avaliada da mesma maneira para todos os participantes?

8. A análise estatística foi apropriada?

Importante, o numerador e o denominador devem ser claramente reportados, e as percentagens devem ser evidenciadas com intervalos de confiança. A secção de métodos

deve ser detalhada o suficiente para que os revisores identifiquem a técnica analítica utilizada e de que forma as variáveis específicas foram avaliadas. Além disto, é igualmente importante avaliar a adequação da estratégia analítica em termos das premissas associadas à abordagem, uma vez que diferentes métodos de análise se baseiam em diferentes pressupostos sobre os dados, e como estes responderão.

9. A taxa de resposta foi adequada e, se não, este número reduzido foi gerido de forma apropriada?

Um grande número de desistências, recusas ou *não encontrados* entre os indivíduos selecionados pode diminuir a validade de um estudo, tal como uma baixa taxa de resposta para os estudos de prevalência. Os autores devem discutir claramente a taxa e quaisquer razões para a não-resposta, comparando os participantes do estudo com os não participantes no estudo, particularmente no que se refere às suas características sociodemográficas. Se os motivos para a não-resposta parecem não estar relacionados com o resultado medido e as características de todos os participantes no estudo forem semelhantes (abordado na questão 5, viés de cobertura/abrangência), os investigadores podem justificar uma taxa de resposta modesta.

2.1.1.1.8 Instrumento de avaliação crítica para estudos de precisão diagnóstica (JBI, 2015b; Whiting et al., 2011).

O instrumento de avaliação crítica para estudos de precisão diagnóstica é composto por 10 itens.

1. A seleção da amostra foi consecutiva ou aleatória?

Os estudos devem indicar ou descrever o método de recrutamento. Se for alegada a escolha de uma amostra aleatória, o método de randomização deve ser indicado (e apropriado). É aceitável se os estudos não disserem *consecutivo* mas, em vez disso, descrevem o recrutamento consecutivo (exemplo: “todos os utentes de ... até ... foram incluídos”).

2. Foi evitado um desenho de caso-controlo?

Os estudos de caso-controlo são descritos detalhadamente no manual de revisores. Em essência, se o desenho de um estudo envolve o recrutamento de participantes já

conhecidos por outros meios de forma a ter o diagnóstico de interesse e investigar se o teste de interesse os identifica corretamente como tal, a resposta é não.

3. O estudo evitou exclusões inadequadas?

Se os participantes forem excluídos por razões que provavelmente influenciariam a conduta, interpretação ou os resultados do teste, isto poderá enviesar os resultados. Exemplos incluem: excluir participantes em que o teste seja de difícil implementação; excluir participantes com resultados limítrofes; excluir participantes com claros indicadores clínicos de diagnóstico de interesse.

4. Foram os resultados do teste diagnóstico de interesse interpretados sem o conhecimento dos resultados do teste de referência?

Os resultados do teste diagnóstico de interesse devem ser interpretados por alguém que é cego aos resultados do teste de referência. O teste de referência pode não ter sido conduzido no mesmo período em que o teste de índice foi realizado e, se assim for, a resposta a esta pergunta será sim. Se a pessoa que interpreta o teste diagnóstico de interesse também interpretou o teste de referência, então presume-se que a esta questão será respondido não, a menos que existam outros fatores em jogo (por exemplo, a interpretação dos resultados pode ser separada da sua colheita, onde o intérprete é cego à identidade do utente e resultados anteriores obtidos em testes de referência).

5. No caso de ser considerado um limiar diagnóstico foi este pré-especificado?

O limiar de diagnóstico pode ser escolhido com base no que permite obter a precisão ótima a partir dos dados ou pode ser pré-especificado. Quando nenhum limiar de diagnóstico é aplicado (isto é, os resultados de um teste baseiam-se na observação de uma característica específica, que pode existir ou não) esta pergunta deve ser respondida como não aplicável.

6. É o teste de referência suscetível de classificar corretamente a condição de interesse?

O teste de referência deve ser o padrão ouro para o diagnóstico da condição de interesse. Adicionalmente, o relatório do estudo deve descrever claramente a sua condução, com detalhes suficientes que possibilitem aos revisores a confiança de que a sua implementação foi realizada de forma correta e competente.

7. Foram os resultados do teste de referência interpretados sem o conhecimento dos resultados do teste diagnóstico de interesse?

Os critérios estabelecidos para a pergunta 4 aplicam-se igualmente aqui. Os resultados do teste de referência devem ser interpretados por alguém que é cego aos resultados do teste diagnóstico de interesse. O teste diagnóstico de interesse pode não ter sido realizado no mesmo período temporal em que o teste de referência é realizado e, se assim for, a resposta a esta pergunta será sim. Se a pessoa que interpreta o teste de referência também interpretou o teste diagnóstico de interesse, então presume-se que a resposta a esta questão será não, a menos que existam outros fatores em jogo (por exemplo, a interpretação dos resultados pode ser realizada separadamente da sua colheita, e neste caso o intérprete pode ser cego no que respeita à identidade do utente e aos resultados do teste diagnóstico de interesse anteriores).

8. Decorreu um intervalo adequado entre a aplicação do teste diagnóstico de interesse e do teste de referência?

O teste diagnóstico de interesse e o teste de referência devem ser realizados o mais próximo possível de modo a que o *status* do utente não possa ter mudado significativamente. O tempo máximo aceitável variará com base nas características da população e condição de interesse.

9. Todos os utentes foram avaliados com o mesmo teste de referência?

O padrão de referência pelo qual os utentes são classificados como tendo, ou não, a condição de interesse deve ser o mesmo para todos os utentes. Se os resultados do teste diagnóstico de interesse influenciarem como ou se o teste de referência é utilizado (por exemplo, onde um falso negativo aparente pode ser detetado, o desenho do estudo pode exigir uma *dupla verificação*), tal pode resultar em estimativas tendenciosas de sensibilidade e especificidade. Além disso, em alguns estudos, dois testes de referência paralelos podem ser utilizados (em utentes diferentes) e os resultados então combinados. Em ambos os casos, os resultados devem ser não.

10. Foram todos os utentes incluídos na análise?

Perdas para o *follow-up* devem ser explicadas, assim como devem ser consideradas se é provável que a sua causa e frequência tenham tido efeito nos resultados (pode existir

subjetividade neste contexto, deve ser mantida baixa tolerância geral na decisão de responder não a esta questão, ainda que uma perda única num estudo coorte de larga escala não deva forçar uma resposta negativa). No entanto, se os resultados de um participante forem difíceis de interpretar e, por isso, excluídos da análise, tal pode resultar na estimativa exagerada de precisão do teste diagnóstico, e esta questão deve ser definitivamente respondida como não.

2.1.1.9 Instrumento de Avaliação Crítica para Estudos Transversais Analíticos (JBI, 2016).

O instrumento de avaliação crítica para estudos transversais analíticos é composto pelos oito itens seguintes.

1. Os critérios para inclusão na amostra foram claramente definidos?

Os autores devem fornecer critérios claros de inclusão e exclusão que eles definiram antes do recrutamento dos participantes do estudo. Os critérios de inclusão/exclusão devem ser especificados (por exemplo: risco, fase de progressão da doença) com detalhes suficientes e toda a informação necessária para o estudo.

2. Os sujeitos do estudo e o contexto foram descritos em detalhe?

A amostra do estudo deve ser descrita com detalhes suficientes para que outros investigadores possam determinar se é comparável à sua população de interesse. Os autores devem fornecer uma descrição clara da população da qual os participantes do estudo foram selecionados ou recrutados, incluindo dados demográficos, localização e período de tempo.

3. A exposição foi medida de forma válida e confiável?

O estudo deve descrever claramente o método de avaliação da exposição. Avaliar a validade requer que haja um padrão-ouro ao qual a medida possa ser comparada. A validade da avaliação da exposição relaciona-se usualmente com a adaptação de uma medida atual ou se uma medida de exposição passada é necessária. Confiabilidade refere-se aos processos incluídos num estudo epidemiológico para verificar a repetibilidade das medições das exposições. Estes geralmente incluem confiabilidade, avaliação intra-observadora e confiabilidade entre observadores.

4. Foram utilizados critérios objetivos, padrão para a medição da condição?

É útil determinar se os utentes foram incluídos no estudo com base num diagnóstico específico ou definição. Tal diminuirá provavelmente o risco de viés. As características são uma outra aproximação útil aos grupos de harmonização, e os estudos que não usaram métodos ou definições específicas de diagnóstico devem fornecer a evidência relativa à combinação por características-chaves.

5. Foram identificados fatores confundentes?

Ocorreu confusão quando o efeito estimado da exposição à intervenção é influenciado pela presença de alguma diferença entre os grupos de comparação (para além da exposição investigada/de interesse). Os fatores confundentes típicos incluem características de linha de base (*baseline*), fatores de prognóstico ou exposições concomitantes (por exemplo, fumar). Um confundente é uma diferença entre os grupos em comparação que influencia a direção dos resultados do estudo. Um estudo de alta qualidade ao nível do desenho da coorte irá identificar os potenciais fatores confundentes e medi-los (sempre que possível). Tal torna-se difícil para estudos onde fatores comportamentais, atitudinais ou de estilo de vida podem afetar os resultados.

6. Foram abordadas estratégias para lidar com fatores confundentes?

Estratégias para lidar com os efeitos de fatores confundentes podem ser tratadas no desenho do estudo ou na análise de dados. Ao combinar ou estratificar a amostragem dos participantes, os efeitos dos fatores confundentes podem ser ajustados. Ao lidar com ajuste na análise de dados devem ser avaliadas as estatísticas utilizadas no estudo. A maioria será alguma forma de análise de regressão multivariada para explicar os fatores confundentes medidos.

7. Os resultados foram medidos de forma válida e confiável?

Leia a secção de métodos do documento. Se, por exemplo, o cancro do pulmão é avaliado com base em definições existentes ou critérios de diagnóstico, então a resposta a esta pergunta é provável que seja *sim*. Se o cancro do pulmão é avaliado utilizando para o efeito relatos observacionais ou escalas autorrelatadas, o risco de excesso ou sub-relato aumenta, e a objetividade é comprometida. Importante, determine se os instrumentos de avaliação utilizados foram instrumentos validados, na medida em que apresentam impacto significativo na validade da avaliação de resultados. Tendo estabelecido a

objetividade do instrumento de avaliação de resultados (por exemplo, cancro do pulmão), é importante estabelecer como foi realizada a avaliação. Os elementos envolvidos na colheita de dados foram treinados ou formados no uso do(s) instrumento(s)? (Por exemplo, radiologistas). Se existiu mais do que um elemento envolvido na colheita de dados, eram estes semelhantes em termos de nível de escolaridade, experiência clínica ou investigacional, ou nível de responsabilidade no trabalho de investigação?

8. Foi utilizada a análise estatística apropriada?

Como qualquer consideração de análise estatística deve ser ponderada a existência de um método estatístico alternativo mais apropriado que poderia ter sido usado. A seção de métodos deve ser suficientemente detalhada para que os revisores identifiquem quais as técnicas analíticas utilizadas (em particular, regressão ou estratificação) e como foram avaliados os fatores confundentes específicos. Para estudos que utilizem análise de regressão, é útil identificar se o estudo reconheceu quais as variáveis incluídas e como estas se relacionam com o resultado. Se a estratificação foi a abordagem analítica utilizada, os estratos de análise foram definidos pelas variáveis especificadas? Além disso, é também importante avaliar a adequação da estratégia analítica em termos dos pressupostos associados à abordagem, uma vez que os diferentes métodos de análise são baseados em diferentes pressupostos sobre os dados e como ele irá responder.

2.1.2 Minimizar o risco de viés numa revisão qualitativa.

Apesar da evidência objetiva e explicativa da realidade, decorrente dos estudos quantitativos e das RSLs quantitativas, ser de extrema importância para informar a prática profissional, há questões que não podem ser respondidas por tais modelos explicativos de investigação.

As RSLs e respetiva metodologia de estudos quantitativos, sobretudo quanto se sintetizam estudos de eficácia, é consensual na comunidade académica. No entanto, o mesmo não pode ser dito relativamente à síntese qualitativa de evidência. Esta abordagem é reconhecida como uma mudança de paradigma de metodologias e de métodos (JBI, 2014a).

Tem-se verificado um aumento do interesse no potencial da investigação qualitativa e a sua síntese pode desempenhar um papel significativo para efetuar mudanças desejadas na saúde, educação e bem-estar social (Sandelowski, Voils, & Barroso, 2006), ou na compreensão de como os indivíduos ou comunidades percebem a saúde ou os processos de saúde-doença (Apóstolo & Gameiro, 2005).

De facto, como apontam Apóstolo e Gameiro (2005), há dimensões do objeto de investigação na área da enfermagem e da saúde que se prestam mais a ser investigadas numa perspetiva empírico-compreensiva com recurso, entre outras, a metodologias qualitativas como a fenomenologia, a etnografia, a teoria fundamentada ou a etnometodologia. Entre essas dimensões podemos considerar, entre outros tópicos, estratégias mais adequadas de educação para a saúde num determinado contexto, adesão a um estilo de vida saudável, experiência de estar internado ou de sofrer um determinado processo de doença, sentido da esperança durante o tratamento oncológico, gestão do quotidiano no confronto com a doença, ou alterações na dinâmica familiar, social ou laboral.

Tal como na investigação quantitativa se justifica a síntese de estudos qualitativos porque os resultados a partir de um único estudo qualitativo não devem ser considerados para produzir recomendações para a prática (JBI, 2014a). Também, tal como na investigação quantitativa, a avaliação crítica dos estudos primários qualitativos é fundamental para estabelecer a sua qualidade (JBI, 2014a; Sandelowski et al., 2006).

Tradicionalmente, os termos usados para avaliar criticamente o rigor da investigação são a confiabilidade (*reliability*) e a validade. Geralmente, a confiabilidade é a medida em que os resultados de um estudo relativamente a uma medição são reprodutíveis em diferentes circunstâncias. Por sua vez, a validade refere-se ao grau em que um estudo reflete ou avalia o conceito específico que o investigador está a tentar medir com precisão. Mas a avaliação crítica de estudos qualitativos apresenta contornos diferentes. Como veremos a seguir, questões de natureza ontológica, epistemológica, metodológica, ética, afiliação, background, experiência e contexto do investigador são fundamentais para aferir o valor da investigação qualitativa primária e podem fazer da síntese dos resultados qualitativos uma tarefa complexa e assustadora (Apóstolo & Gameiro, 2005; JBI, 2014a; Sandelowski et al., 2006).

Tal como a confiabilidade/reprodutibilidade das medidas, a validade interna e externa e a objetividade são critérios considerados essenciais para a qualidade de estudos quantitativos, nos estudos qualitativos devemos considerar critérios correspondentes. Neste sentido, os conceitos de *dependability*, *credibility*, *transferability* e *confirmability*, apontados por Lincoln e Guba (1985), deverão ser considerados para estabelecer o valor de um estudo qualitativo, como desenvolveremos de seguida.

Dependability, que traduzimos como confiabilidade, mostra a consistência e congruência dos resultados. Na investigação qualitativa o objetivo da confiabilidade não está em obter exatamente o mesmo resultado uma e outra vez, como acontece na investigação quantitativa, mas sim na obtenção de similaridade e consistência dos resultados.

Para que se possa avaliar a confiabilidade, o processo metodológico de cada estudo deve ser detalhadamente descrito, permitindo assim que outro investigador possa repetir o

mesmo estudo, não necessariamente para obter os mesmos resultados (Shenton, 2004). Shenton (2004) e Anney (2014) resumiam alguns procedimentos que podem ser feitos pelos investigadores: registo de todos os passos e atividades de investigação para mostrar como os dados foram colhidos, registados e analisados; replicação passo a passo, em que dois ou mais investigadores analisam os mesmos dados separadamente e comparam os resultados; estratégias de codificação em que cada investigador recodifica os mesmos dados uma segunda vez decorrido um período de uma ou duas semanas após a primeira codificação; discussão do processo e dos resultados com pares neutros ao processo.

Como veremos, quando nos debruçarmos sobre o instrumento de avaliação crítica de estudos qualitativos (JBI, 2014a), a consistência entre a ontologia, a epistemologia, a metodologia e os métodos são aspetos fundamentais para aferir a consistência e a congruência dos resultados.

Credibility, que traduzimos como credibilidade, é um dos critérios fundamentais para estabelecer a veracidade (*trustworthiness*) dos achados, ou seja, o quão congruentes são os achados com a realidade (corresponde à validade interna nos estudos quantitativos). Pode ser estabelecida a partir da perspetiva do participante na investigação. Envolve o estabelecimento de convicção de que os achados têm fundamento no discurso dos participantes ou observação de campo realizada.

Para promover a confiança e precisão do fenómeno sob escrutínio, Shenton (2004) e Anney (2014) consideram alguns procedimentos que passamos a resumir: adoção de métodos de investigação bem estabelecidos; desenvolvimento, por parte do investigador, de uma familiaridade precoce com a cultura das organizações e dos participantes (por exemplo através da emersão no contexto); triangulação; utilização de técnicas para ajudar a garantir a honestidade dos informantes; análise de caso negativo; interação frequente entre o investigador e supervisores/orientadores, habitualmente mais experientes, relativamente ao plano e desenvolvimento da investigação; reflexão individual ao longo do desenvolvimento do projeto; experiência e qualificações do investigador.

De acordo com os critérios do JBI (2014a, p. 40), os achados podem ser considerados *inequívocos* (achados acompanhados por informação, por exemplo, por uma ilustração textual do discurso do[s] participante[s] ou observações de campo não passível de contestação), *credíveis* (resultados acompanhados por dados, ou seja, por uma ilustração textual do discurso do[s] participante[s] ou observações de campo, mas com falta de clara associação entre os dois e, portanto, passível de possível impugnação), ou *não suportados* (achados não apoiados por informação credível).

Transferability, que traduzimos como transferibilidade (corresponde à validade externa ou generalização nos estudos quantitativos), mostra que os resultados têm aplicabilidade

noutros contextos semelhantes. A discussão sobre a generalização/transferibilidade dos resultados de estudos qualitativos mantém-se acesa no seio da comunidade científica.

Anney (2014) e Shenton (2004) consideraram um conjunto de pressupostos a ter em conta antes de qualquer tentativa de transferência dos resultados de estudos qualitativos para contexto idêntico, entre outras: o número e características das organizações que participaram no estudo; limitações no tipo de pessoas que contribuíram para os dados; amostragem intencional dos participantes; descrição detalhada das características dos participantes envolvidos no trabalho de campo; métodos de recolha de dados que foram utilizados; número e duração das sessões de recolha de dados; período de tempo durante o qual os dados foram colhidos; área geográfica em que o trabalho de campo foi realizado.

Confirmability, que traduzimos como confirmabilidade (correspondente à objetividade nos estudos quantitativos), mostra o grau de neutralidade do investigador e dos processos de análise, ou seja, em que medida os resultados de um estudo correspondem à informação obtida a partir dos respondentes e não foram enviesados por qualquer motivação ou interesse do investigador. Uma auditoria feita por outros investigadores e triangulação reflexiva podem ser passos importantes para controlar este aspeto, como já descrito na confiabilidade (Anney, 2014; Shenton, 2004).

2.1.2.1 Avaliação crítica e validade dos estudos qualitativos.

Havendo uma enorme quantidade de literatura científica qualitativa disponível o objetivo é incluir apenas os estudos que os revisores consideram ter padrões aceitáveis. A avaliação crítica dos artigos identificados e que correspondem aos critérios de inclusão é fundamental no processo de realização de uma síntese qualitativa utilizando ou não meta-agregação.

Seguidamente, descreveremos os itens e os critérios do instrumento de avaliação crítica de estudos qualitativos e dos textos opinião que adaptamos para português (JBI, 2014a). Tal como nos instrumentos anteriormente apresentados para a investigação quantitativa, os revisores, em cada item tem quatro opções que devem considerar: *sim, não, não claro* ou *não aplicável*.

2.1.2.1.1 Instrumento de avaliação crítica para estudos qualitativos (JBI, 2014a; Lockwood, Munn, & Porritt, 2015).

Este instrumento é constituído por 10 itens avaliando a congruência onto-epistemológica, metodológica e ética, e o posicionamento do investigador face à investigação. De notar que esta grelha não contempla os aspetos de transferibilidade anteriormente discutidos.

1. Congruência entre a perspectiva filosófica declarada e a metodologia da investigação

O relatório indica claramente as premissas filosóficas ou teóricas em que o estudo se baseia? O relatório indica claramente a abordagem metodológica adotada em que o estudo se baseia? Existe congruência entre os dois? Por exemplo:

Um relatório pode indicar que o estudo adotou uma perspectiva crítica e uma metodologia participativa de investigação-ação. Aqui há congruência entre a visão crítica (focando no conhecimento decorrente da crítica, ação e reflexão) e investigação-ação (uma abordagem que se concentra primeiramente no trabalho com grupos de modo a refletir sobre questões ou práticas e, em seguida, considera como podem ser diferentes; então age para criar a mudança e, finalmente, identifica novos conhecimentos decorrentes da ação tomada). No entanto, um relatório pode indicar que o estudo adotou uma perspectiva interpretativa e utilizou questionários. Neste sentido verifica-se a incongruência entre a visão interpretativa (centrada no conhecimento decorrente do estudo dos significados atribuídos por indivíduos e grupos a determinado fenómeno) e questionários (uma abordagem centrada em perguntas-padrão a uma população de estudo definida); um relatório pode indicar que o estudo foi qualitativo ou que foi utilizada metodologia qualitativa (tais declarações não demonstram rigor no desenho) ou não fazem nenhuma declaração sobre orientação ou metodologia filosófica.

2. Congruência entre a metodologia e a questão ou objetivos da investigação

A metodologia de estudo é adequada na abordagem da questão de investigação? Por exemplo: num estudo pode ser afirmado que a questão de investigação reside na procura do significado da dor num grupo de pessoas com determinada condição clínica e que foi adotada uma abordagem fenomenológica. Neste caso, há congruência entre essa questão e a metodologia. No entanto, um relatório pode indicar que a questão de investigação foi determinar os efeitos do aconselhamento sobre a severidade da experiência de dor tendo sido seguida uma abordagem etnográfica. Uma pergunta que procura estabelecer a causa e o efeito não pode ser respondida usando uma abordagem etnográfica (dado que a etnografia se propõe a desenvolver entendimentos sobre práticas culturais), demonstrando clara incongruência.

3. Congruência entre a metodologia da investigação e os métodos usados para colher os dados

Os métodos de colheita de dados são apropriados para a metodologia? Por exemplo:

Um relatório pode indicar que o estudo prosseguiu uma abordagem fenomenológica e os dados foram colhidos através de entrevistas fenomenológicas. Existe congruência entre a metodologia e a colheita de dados. Um relatório pode indicar que o estudo

prosseguiu uma abordagem fenomenológica e os dados foram colhidos através de um questionário enviado por correio. Existe incongruência entre a metodologia e a colheita de dados, uma vez que a fenomenologia procura obter descrições ricas da experiência de um fenómeno que não pode ser alcançado através da procura de respostas escritas a questões padronizadas. Existe congruência entre a metodologia de pesquisa e a representação e análise de dados.

4. Congruência entre a metodologia da investigação e a representação e análise de dados

Os dados foram analisados e representados de forma a serem congruentes com a posição metodológica declarada? Por exemplo:

Um relatório pode indicar que o estudo prosseguiu uma abordagem fenomenológica para explorar a experiência das pessoas sobre o sofrimento, pedindo aos participantes que descrevem as suas experiências de sofrimento. Se o texto elaborado a partir destas questões procura estabelecer o significado de sofrimento para os participantes, e os significados de todos os participantes estão incluídos nos resultados do relatório, então verifica-se a existência de congruência; o mesmo relatório pode, no entanto, concentrar-se apenas nos significados que eram comuns a todos os participantes e descartar significados isolados relatados. Tal não seria apropriado no trabalho fenomenológico.

5. Congruência entre a metodologia de pesquisa e interpretação dos resultados

Os resultados são interpretados de forma apropriada à metodologia? Por exemplo:

Um relatório pode indicar que o estudo prosseguiu uma abordagem fenomenológica para explorar a experiência das pessoas relativamente à desfiguração facial e os resultados são utilizados para informar os profissionais sobre personalizar os cuidados prestados de acordo com as diferenças individuais. Existe congruência entre a metodologia e a abordagem de interpretação; um relatório pode indicar que o estudo prosseguiu uma abordagem fenomenológica para explorar a experiência das pessoas relativamente à desfiguração facial e os resultados são utilizados para criar listas de verificação práticas para contextos de avaliação. Existe incongruência entre a metodologia e a abordagem de interpretação, na medida em que a fenomenologia procura entender o significado de um fenómeno para os participantes do estudo, a qual não pode ser interpretada de forma generalizada para o total da população a um nível em que as avaliações padronizadas sejam relevantes para todos os indivíduos.

6. Posicionamento do investigador cultural ou teoricamente.

São as crenças e valores, e a sua potencial influência, declarados no estudo? Por exemplo:

O investigador desempenha um papel substancial no processo de investigação qualitativa e é importante, na avaliação das evidências decorrentes, conhecer a orientação cultural e teórica do investigador. Um estudo de alta qualidade incluirá um esclarecimento deste ponto.

7. Influência do investigador sobre a pesquisa e vice-versa.

O potencial de influência do investigador no estudo e o potencial de influência do próprio processo de investigação no investigador, e nas interpretações que faz, são reconhecidas e abordadas? Por exemplo:

A relação entre o investigador e os participantes do estudo é abordada? O investigador analisa criticamente o seu papel e a sua potencial influência durante a colheita de dados? É reportado como o investigador respondeu a eventos que emergiram durante o estudo?

8. Representação dos participantes e das suas vozes.

Usualmente, os estudos devem fornecer ilustrações dos dados de modo a sustentar a base das suas conclusões e garantir que os participantes estejam representados. A representação dos participantes e das suas vozes envolve o estabelecimento de convicção de que os achados têm fundamento no discurso dos participantes ou observação de campo realizada. Os participantes e os excertos do seu discurso deverão estar devidamente representados no relatório para que se possa estabelecer a credibilidade dos achados.

9. Aprovação por comité de ética.

Uma declaração sobre o processo de aprovação ética deve ser feita no estudo.

10. Relação das conclusões com a análise e interpretação dos dados.

Este critério diz respeito à relação entre os achados reportados e as opiniões ou palavras dos participantes em estudo. Na avaliação de um estudo, os revisores devem certificar-se de que as conclusões extraídas na investigação são baseadas nos dados colhidos; dados são produto textual obtido através de observação, entrevistas ou outros processos.

2.1.2.1.2 Instrumento de Avaliação Crítica para Estudos de Texto e Opinião (McArthur, Klugarova, Yan, & Florescu, 2015).

O instrumento de avaliação crítica para estudos de texto e opinião é composto pelos seis itens seguintes.

1. A origem da opinião está claramente identificada?

Existe referência ao autor? Peças editoriais sem nome em revistas ou jornais, potencializam a sua análise e comentário, no entanto a autoria deve ser identificada.

2. A fonte da opinião é uma referência na área?

As qualificações, posição e afiliações atuais do autor com grupos específicos devem ser indicadas na publicação, e o revisor deverá estar satisfeito com as competências reconhecidas do(s) autor(es) no campo científico.

3. Os interesses da população em causa são o ponto central do parecer?

O objetivo desta pergunta é estabelecer o objetivo do autor ao escrever o artigo, considerando o público-alvo pretendido. Se o tópico da revisão está relacionado com uma intervenção clínica, ou com um aspeto da prestação de cuidados de saúde, o foco nos resultados em saúde torna-se pertinente. No entanto, se o intuito do estudo estiver relacionado com o comportamento interpessoal ou relações de poder, é desejável o enfoque nos grupos relevantes. Por conseguinte, esta questão deve ser respondida no contexto da finalidade da revisão.

4. A posição declarada é o resultado de um processo analítico, e há lógica na opinião expressa?

A fim de estabelecer a clareza da fundamentação ou base do parecer, analise a orientação das principais linhas de argumentação. Questões a colocar na análise de cada estudo de texto e opinião: Quais são os principais pontos nas conclusões ou recomendações? O autor usa que argumentos para os suportar? O argumento é lógico? Foram claramente definidos termos importantes? Os argumentos suportam os pontos principais?

5. Há referência à literatura existente?

Se subsistir referência à literatura existente, esta representação é não-tendenciosa, inclusiva, ou será uma descrição não-crítica do conteúdo que apoia especificamente a linha de argumentação apresentada? Estas considerações enfatizam a robustez de como a literatura citada foi utilizada.

6. Qualquer incongruência com a literatura/fontes é logicamente defendida?

É feita referência no texto se o parecer expresso apresenta suporte mais amplo? Considere também se o autor evidenciou outras opiniões alternativas ou dominantes na literatura e indicou/defendeu a sua posição de forma informada, relacionando-a com outros discursos ou discursos similares.

2.1.3 Minimizar o risco de viés numa revisão de revisões.

Tal como acontece em revisões sistemáticas que incluem estudos primários (quantitativos ou qualitativos), uma revisão de revisões deverá também incluir a avaliação crítica das RSLs que cumprem os critérios de inclusão.

2.1.3.1 Instrumento de Avaliação Crítica para Revisões Sistemáticas e Síntese de Investigação (Aromataris et al., 2015; JBI, 2014d, pp. 30-32).

O instrumento de avaliação crítica para estudos de RSL e síntese de investigação deve ser usado quando se desenvolvem revisões de revisões. É composta pelos 11 itens seguintes.

1. A questão de revisão está enunciada de forma clara e explícita?

A questão da revisão é um passo essencial no processo de revisão sistemática. Uma questão bem articulada define o âmbito da revisão e assiste no desenvolvimento da estratégia de pesquisa de modo a localizar a evidência relevante. Uma questão explicitamente declarada, formulada em torno dos seus elementos PICO (População, Intervenção, Comparador, Resultado), ou respetiva adaptação para cada revisão, assiste

a equipa de revisão na condução da revisão e assiste o leitor a determinar se a revisão alcançou os seus objetivos. Idealmente, a questão de revisão deve ser articulada num protocolo publicado; todavia, tal nem sempre será o caso.

2. Os critérios de inclusão foram apropriados para a questão de revisão?

Os critérios de inclusão devem ser identificáveis e coincidir com a questão de revisão. Os elementos necessários do PICO, ou respetiva adaptação para cada revisão, devem ser explícitos e claramente definidos. Os critérios de inclusão devem ser detalhados e as revisões incluídas devem ser claramente elegíveis quando comparadas com os critérios de inclusão estabelecidos. Os avaliadores de meta-análises deparar-se-ão com critérios de inclusão que podem incluir a capacidade de realizar análises estatísticas que não seriam a norma para uma revisão sistemática. Os tipos de estudos incluídos devem ser relevantes para a questão da revisão, por exemplo, uma revisão *umbrella* que visa resumir uma série de intervenções não-farmacológicas eficazes em comportamentos agressivos entre utentes idosos com demência. Esta revisão limitar-se-á a incluir análises sistemáticas e meta-análises que sintetizem estudos quantitativos de modo a avaliar as diversas intervenções; revisões qualitativas ou económicas não seriam incluídas.

3. A estratégia de pesquisa foi adequada?

Uma revisão sistemática deve fornecer evidências da estratégia de pesquisa utilizada para localizar a evidência. Tal pode ser encontrado, em alguns casos, na seção de métodos da revisão ou enquanto apêndice que pode ser fornecido como informação complementar à publicação da revisão. Uma revisão sistemática deve apresentar uma estratégia de pesquisa clara que aborda cada um dos componentes PICO (ou respetiva adaptação para cada revisão) identificáveis da questão de revisão. Algumas revisões também podem fornecer uma descrição da abordagem de pesquisa e de onde derivaram os termos utilizados, embora devido aos limites de palavras em revistas isto possa constituir a norma somente em publicações on-line. Deverá existir evidência relativa às palavras-chave e termos lógicos e relevantes, assim como para *Subject Headings* e termos de indexação utilizados na condução da pesquisa. Os limites da pesquisa e o seu potencial impacto também devem ser considerados; por exemplo, se um limite temporal foi utilizado, este foi apropriado e/ou justificado? Se apenas estudos de língua inglesa foram incluídos, terá tal viés de linguagem impacto na revisão? A resposta a estas considerações dependerá, em parte, da questão de revisão.

4. As fontes e os recursos utilizados para procurar os estudos foram adequados?

Uma revisão sistemática deve tentar identificar toda a evidência disponível e, como tal, deve existir evidência de uma estratégia abrangente de pesquisa. Várias bases de dados eletrónicas devem ser utilizadas, incluindo grandes bases de dados bibliográficas como a MEDLINE e CINAHL. Idealmente, outras bases de dados relevantes para a questão de revisão devem também ser pesquisadas, por exemplo, uma revisão sistemática com uma questão sobre intervenções de fisioterapia também deve pesquisar na base de dados *Physiotherapy Evidence Database* (PEDro), enquanto numa revisão focada em intervenções educacionais deve ser utilizada a *Education Resources Information Center* (ERIC). Revisões de eficácia devem pesquisar em locais de registo de ensaios clínicos. Uma pesquisa abrangente é a forma ideal de minimizar o viés de publicação, como resultado, uma revisão sistemática bem conduzida também deverá tentar procurar literatura cinzenta ou estudos não publicados; tal processo poderá envolver a pesquisa em websites relevantes para a questão de revisão ou repositórios de teses.

5. Os critérios de avaliação crítica dos estudos foram apropriados?

Uma revisão sistemática deve apresentar uma declaração clara de que a avaliação crítica foi realizada e fornecer os detalhes dos itens que foram utilizados para avaliar os estudos incluídos. Tal descrição pode ser apresentada nos métodos da revisão, como um apêndice de informações complementares, ou enquanto referência a uma fonte localizável. As ferramentas ou instrumentos utilizados devem ser apropriados para a questão de revisão e o tipo de investigação realizada. Por exemplo, uma revisão sistemática de efetividade deve apresentar uma ferramenta ou instrumento que aborde aspetos de validade para estudos experimentais e ensaios clínicos randomizados, como aleatorização e ocultação - se a revisão incluir investigação observacional de modo a responder à mesma pergunta, uma ferramenta diferente será mais apropriada. Da mesma forma, uma revisão que avalie a precisão do teste de diagnóstico pode referir-se à ferramenta reconhecida QUADAS (Whiting, Rutjes, Reitsma, Bossuyt, & Kleijnen, 2003).

6. A avaliação crítica foi realizada por dois ou mais revisores de forma independente?

A avaliação crítica ou a avaliação similar da qualidade da literatura incluída numa revisão sistemática é essencial. Uma característica-chave para minimizar o viés ou erro sistemático na condução de uma revisão sistemática é a realização da avaliação crítica dos estudos incluídos de forma independente e em duplicado por membros da

equipa de investigação. A revisão sistemática deve apresentar uma declaração clara de que a avaliação crítica foi realizada por, pelo menos, dois revisores que trabalharam independentemente e conferenciaram, quando necessário, de modo a decidirem sobre a qualidade do estudo e a elegibilidade com base na qualidade.

7. Foram considerados procedimentos para minimizar erros na extração de dados?

Os esforços realizados pelos autores da revisão durante a extração de dados também podem minimizar o viés ou erros sistemáticos na condução de uma revisão sistemática. As estratégias para minimizar o viés podem incluir a realização de toda a extração de dados em duplicado e de forma independente, utilizando ferramentas ou instrumentos específicos para orientar a extração de dados e algumas informações relativas ao treino no seu uso.

8. Os métodos utilizados para combinar os estudos foram apropriados?

A síntese da evidência é uma característica fundamental de uma revisão sistemática. A síntese que é apresentada deve ser apropriada para a questão de revisão, o tipo de revisão sistemática e evidência a que se refere. Se foi realizada meta-análise, esta necessita de ser cuidadosamente revista. A combinação de estudos foi apropriada? Os revisores avaliaram estatisticamente a heterogeneidade e forneceram alguma explicação para a heterogeneidade que possa estar presente? Frequentemente, quando estudos heterogêneos estão incluídos na revisão sistemática, a síntese narrativa será um método adequado para a apresentação dos resultados de múltiplos estudos. Se for uma revisão qualitativa, os métodos que foram utilizados para sintetizar os achados são congruentes com a metodologia da revisão? Existe informação descritiva e explicativa adequada que apoie que os achados sintetizados finais foram construídos a partir dos achados provenientes da investigação original?

9. A probabilidade de viés de publicação foi avaliada?

Como mencionado, uma estratégia de pesquisa abrangente é o melhor meio pelo qual o autor duma revisão pode minimizar o impacto do viés de publicação nos resultados da revisão. As revisões devem, para além disso, apresentar testes estatísticos, como o teste de funil de *Egger* para avaliar também a presença potencial de viés de publicação e o seu potencial impacto sobre os resultados da revisão. Esta questão não será aplicável em revisões sistemáticas de evidências qualitativas.

10. As recomendações para a prática e/ou política são apoiadas pelos resultados divulgados?

Embora as primeiras nove questões visem especificamente o viés potencial na condução de uma revisão sistemática, as questões finais são mais indicadoras da qualidade do que da validade da revisão. Idealmente, uma revisão deve apresentar recomendações para a prática e/ou política. Quando estas recomendações são feitas deve haver uma clara relação entre as referidas recomendações e os resultados da revisão. Existe clara evidência de que a robustez/fragilidade dos resultados e a qualidade da revisão foram considerados na formulação das recomendações da revisão?

11. As orientações para o desenvolvimento de investigação futura são adequadas?

O processo de revisão sistemática é reconhecido pela sua capacidade de identificar a existência de lacunas na investigação, ou base de conhecimento, em torno de um determinado tópico. A maioria dos autores de revisões sistemáticas fornecerá alguma indicação, frequentemente na discussão, relativa à direção futura da investigação. Quando a evidência é escassa ou o tamanho das amostras que suportam as estimativas globais do efeito são pequenas, e as estimativas do efeito são imprecisas, a repetição de investigações semelhantes às identificadas pela revisão pode ser necessária e apropriada. Em outras situações, pode ser justificada a necessidade de novas questões de investigação para investigar o tópico.

CAPÍTULO III

TIPOS DE REVISÃO SISTEMÁTICA DA LITERATURA

Para além da síntese de evidência de estudos quantitativos (síntese de evidência de eficácia) e qualitativos (síntese de evidência de significado) ou de estudos de análise económica de uma determinada intervenção, processo ou procedimento, pode ser desenvolvida uma gama de metodologias relacionadas com as tipologias e desenhos de estudos primários e com as perguntas de revisão.

Nos últimos anos, o JBI, dentro da sua colaboração, constituiu um conjunto de grupos, denominados como grupos metodológicos, que têm desenvolvido os pressupostos, os métodos e os manuais, acessíveis na sua página, relativamente a um conjunto de novas metodologias de revisão sistemática. Entre elas encontram-se: revisões de revisões (*umbrella reviews*), revisões de prevalência e incidência (*prevalence and incidence reviews*), revisões de precisão do diagnóstico (*diagnostic test accuracy reviews*), revisões com métodos mistos (*mixed methods reviews*), revisões *scoping* (*scoping reviews*) e revisões de avaliação económica (*reviews of economic evaluations*). Relativamente à avaliação económica, as respetivas RSLs podem ser desenvolvidas em relação ao custo minimização, custo efetividade, custo benefício ou custo utilidade (JBI, 2014a). Tendo em conta estas complexidades e a necessidade de especialização nesta área, não iremos, neste documento, proceder à sua abordagem.

Tal como acontece com qualquer investigação é importante ter uma questão de revisão clara e respondível. Uma questão de revisão clara auxilia os leitores a tomar a decisão de ler ou não a publicação e facilita a indexação em bases de dados on-line, tais como MEDLINE ou CINHALL. Além disso, uma boa, clara e precisa questão orienta todo o processo de revisão, quer no que respeita aos critérios de inclusão, estratégia de pesquisa, seleção dos estudos ou síntese da informação relevante para lhe dar resposta. Pelo contrário, a incapacidade de desenvolver uma pergunta robusta vai aumentar a probabilidade de uma revisão pobre e não focada.

Para verificar se existe já um processo de revisão em desenvolvimento, que responda ao mesmo questionamento, aconselha-se uma pesquisa nas bases de dados maioritariamente aceites pela comunidade científica, como a PUBMED/MEDLINE, CINAHL, JBI Connect+, Cochrane Library ou Prospero, antes de se proceder ao desenvolvimento de uma questão de revisão. Tendo em conta a especificidade da revisão, pode ainda ser consultada uma base de dados de uma área específica, por exemplo, que dissemine resultados económicos.

O uso de mnemónicas pode ajudar a clarificar a estrutura de títulos e de questões de revisão. Habitualmente, as questões de revisão numa RSL são orientadas pela mnemónica PICO (População, Intervenção/interesse ou exposição, Controlo ou comparador e resultados a serem medidos/*Outcome*) ou respetivo ajustamento que depende do tipo de revisão. As revisões podem ser orientadas, inicialmente, por um objetivo que depois é operacionalizado com um conjunto de questões mais específicas às quais será mais fácil dar resposta.

As revisões Cochrane têm sido fundamentalmente orientadas por perguntas sobre os efeitos de intervenções e têm-se concentrado principalmente em ensaios clínicos randomizados. No entanto, há uma gama de mnemónicas específicas para orientar a estruturação de questões de revisão de cada tipologia metodológica, tal como veremos já de seguida.

3.1 RSL de evidência de eficácia

O objetivo global de revisões de evidência de eficácia é informar sobre a eficácia de uma intervenção num resultado ou resultados/*outcomes* (JBI, 2014a).

Título

A mnemónica PICO deve ser usada para construir um título claro de uma revisão sistemática de evidência de eficácia, traduzindo População, Intervenção, Comparador (se for caso disso), Resultado (*Outcome*) e a eficácia procurada (JBI, 2014a).

Exemplo: *The effectiveness of nonpharmacological interventions in older adults with depressive disorders: a systematic review* (Apóstolo, Bobrowicz-Campos, Rodrigues, Castro, & Cardoso, 2016; Apóstolo, Queirós, Rodrigues, Castro, & Cardoso, 2015). Os exemplos que se seguem são baseados nestas referências.

Questão de revisão

A questão de revisão em estudos de evidência de eficácia segue a mnemónica **PICO**: População, Intervenção, Comparador e Resultado (*Outcome*).

Exemplo: No exemplo que se segue apresentamos um objetivo geral que é operacionalizado em diferentes questões de revisão.

O objetivo deste estudo foi rever sistematicamente a eficácia de intervenções não farmacológicas em idosos com perturbação depressiva. Mais especificamente, a revisão incidiu sobre as seguintes questões:

I - São as intervenções não farmacológicas eficazes na redução dos sintomas depressivos nos idosos?

II - A eficácia das intervenções varia em função da gravidade dos sintomas depressivos?

III - É a eficácia das intervenções influenciada pela presença ou ausência de um tratamento farmacológico antidepressivo?

IV - É a eficácia das intervenções influenciada pela presença de comorbidades?

População

As características mais importantes na população incluem os seus fatores demográficos (por exemplo, idade, sexo, etnia), fatores socioeconómicos e o contexto (por exemplo, hospital, comunidade, *etc.*)

Exemplo: Esta revisão considera estudos que incluíram pessoas idosas, com idade superior a 65 anos, com qualquer tipo de perturbação depressiva, independentemente das comorbidades e quaisquer tratamentos anteriores. Foram excluídos aqueles com episódios/sintomas maníacos ou psicóticos. Os utentes que recebem tratamento farmacológico para depressão ou outras doenças foram incluídos.

Intervenção e Comparador

A intervenção primária de interesse corresponde, habitualmente, àquela a que foi submetido o grupo de tratamento. O comparador corresponde, habitualmente, à intervenção a que foi submetido o grupo de controlo. O comparador pode ser passivo (placebo, nenhum tratamento, cuidados padrão ou um controlo de lista de espera, por exemplo a intervenção reminiscência vs o comparador cuidados usuais) ou ativo (variação da intervenção, uma droga ou tipo de terapia, por exemplo a intervenção soro fisiológico vs o comparador água).

Exemplo: Nesta revisão foram considerados estudos sobre a eficácia de intervenções não farmacológicas em idosos com perturbação depressiva. Essas intervenções podem incluir, mas não estão limitadas a, terapia com música, massagem, terapia de reminiscência ou exercício físico com qualquer frequência, intensidade ou duração. Foram considerados estudos que compararam as intervenções não farmacológicas com as intervenções de controlo, como o tratamento habitual ou outro comparador ativo.

Resultados

Os resultados podem incluir sobrevivência (mortalidade), eventos clínicos (AVC ou enfarte do miocárdio), resultados relatados pelo utente (por exemplo, qualidade de vida), eventos adversos, sobrecarga do cuidador, restrições ao seu estilo de vida, etc. Podem ter uma medição direta e objetiva (ex: pressão arterial, número de AVCs) ou indireta pelo relato dos indivíduos avaliados (escalas de avaliação ou grelhas de observação) ou, ainda, por meio de avaliação clínica com os critérios definidos (JBI, 2014a).

Exemplo: Como resultados primários, a revisão considerou estudos que incluíam os seguintes: sintomatologia depressiva, medida por qualquer escala de depressão (como a Escala de Depressão Geriátrica) ou mudanças na gravidade dos sintomas depressivos, avaliadas de acordo com critérios de diagnóstico (como a Classificação Internacional de Doenças-10 ou o Manual de Diagnóstico e Estatístico das Perturbações Mentais-IV-Texto Revisto).

Como resultados secundários foram considerados o nível de autonomia nas atividades de vida diária, função cognitiva, qualidade de vida, medidos por instrumentos validados (como o Índice de Lawton-Brody de Atividades Instrumentais de Vida Diária, o Mini-Exame do Estado Mental, Avaliação Cognitiva de Montreal, ou a Escala de Qualidade de Vida da Organização Mundial da Saúde, versão de 100 itens).

Tipo de estudos

O desenho dos estudos a incluir deverá ser específico para dar resposta à questão de revisão. Nas revisões de eficácia são habitualmente incluídos estudos de natureza experimental podendo também ser considerados estudos observacionais.

Exemplo: Nesta revisão considerou-se qualquer desenho do tipo experimental, incluindo estudos randomizados controlados, não randomizados ou outros estudos quase-experimentais, incluindo estudos antes-após.

Síntese dos dados

Neste tipo de revisões, os dados quantitativos podem ser apresentados em tabelas e analisados de forma narrativa e, sempre que possível, podem ser agrupados numa meta-análise. Os tamanhos de efeito podem ser expressos, por exemplo, em *odds ratio* ou como risco relativo (para dados categóricos) e diferenças de média ponderada (para dados contínuos), bem como os respetivos intervalos de confiança (IC) a 95% de confiança. A heterogeneidade deverá ser avaliada do ponto de vista clínico, metodológico e estatístico. A plausível explicação para as variações nos efeitos do tratamento pode ser explorada através de análises de subgrupos, sempre que possível, com base, por exemplo, em diferentes subtipos de classificação clínica dos participantes incluídos nos estudos. Exemplo destes subtipos podem ser os níveis de perturbação depressiva dos participantes, ou diferentes contextos, tais como amostras de sujeitos analisados em contexto hospitalar vs comunidade (Apóstolo, Bobrowicz-Campos et al., 2016; JBI, 2014a; Apóstolo, Cooke et al., 2016, in press).

3.1.1 Meta-análise

A meta-análise é uma parte integrante da RSL. Os resultados de estudos individuais semelhantes são combinados para determinar o efeito global de uma intervenção em relação a outra intervenção de controlo para um determinado resultado. Na meta-análise é calculado o tamanho do efeito e o peso de cada estudo. O tamanho do efeito indica a direção e magnitude dos resultados de um estudo particular. O peso é indicativo da quantidade de informação que um estudo fornece para a análise global no contexto da combinação dos estudos (JBI, 2014a; Santos & Cunha, 2013; Tufanaru et al., 2015).

No entanto há um conjunto de pressupostos a que deve obedecer uma meta-análise. A meta-análise é apropriada somente quando os estudos são considerados semelhantes/homogéneos do ponto de vista clínico (variabilidade nas populações estudadas, nas intervenções envolvidas e nos resultados medidos), metodológico (desenho do estudo e risco de viés – já descrito) e estatístico (variabilidade nos efeitos observados do tratamento que está além do que seria de esperar por erro aleatório¹). A heterogeneidade clínica está intimamente ligada à heterogeneidade estatística, especialmente porque a ocorrência de heterogeneidade clínica pode levar a heterogeneidade estatística (Tufanaru et al., 2015; West et al., 2010).

1. Esta é uma definição aceite por: AHRQ, Centre for Reviews and Dissemination (CRD), a Cochrane Collaboration, o Drug Effectiveness Review Project e a European Network for Health Technology Assessment (West et al., 2010).

O julgamento relativo à homogeneidade/heterogeneidade dos estudos e da adequabilidade da sua combinação estatística deve ser baseado na compreensão da questão de revisão, nas características dos estudos, bem como na facilidade de interpretação dos resultados, e não apenas na heterogeneidade estatística (Santos & Cunha, 2013; Tufanaru et al., 2015).

A análise do pressuposto da heterogeneidade estatística deve iniciar-se por uma exploração visual do gráfico de meta-análise (*forest plot*), avaliando se os estudos individuais mostram uma direção e magnitude do efeito semelhante e se os intervalos de confiança se sobrepõem (Higgins & Green, 2011; JBI 2014a; Thompson & Higgins, 2002). A presença de heterogeneidade estatística deve, para além do anteriormente referido, ser avaliada pela análise do teste *Q* de Cochran. Para a quantificar deve ser usado o índice de I^2 , ou, ainda por análise de meta-regressão, técnicas que abordaremos de seguida (Higgins, Thompson, Deeks, & Altman, 2003; Thompson & Higgins, 2002).

A hipótese nula relativa à heterogeneidade é: dada uma probabilidade aceitável de erro de tipo I, o efeito do tratamento não difere tendo em conta diferentes fatores como a idade, o sexo, a gravidade da doença, o contexto ou a qualidade do estudo (West et al., 2010).

Igualmente, a heterogeneidade clínica tem a mesma hipótese subjacente, mas não inclui questões de desenho dos estudos (metodológicas). Além disso, a heterogeneidade estatística é mais ampla do que heterogeneidade clínica na medida em que abrange heterogeneidade clínica e metodológica, bem como o erro aleatório (West et al., 2010).

Para quantificar a heterogeneidade foi desenvolvida a estatística I^2 que proporciona uma medida do grau de inconsistência nos resultados dos estudos. O I^2 descreve o percentual da variação total entre os estudos que se deve à heterogeneidade e não ao acaso. Pode ser prontamente calculado pela seguinte fórmula: $I^2 = 100\% \times (Q - GL)/Q$; em que *Q* é estatística de heterogeneidade de Cochran e *GL* os graus de liberdade. Os valores negativos de I^2 são igualados a zero pelo que o I^2 pode variar entre 0% e 100%. Um valor de 0% indica que não há heterogeneidade observada, e valores mais elevados mostram heterogeneidade crescente. Um valor próximo de 25% indica baixa heterogeneidade, próximo a 50% indica heterogeneidade moderada e próximo de 75% indica alta heterogeneidade entre os estudos (Higgins et al., 2003).

É ainda conveniente usar a meta-regressão para explorar fontes de heterogeneidade mesmo quando um teste global inicial para a heterogeneidade não for significativo. A meta-regressão investiga se determinadas covariáveis (potenciais modificadores de efeito, por exemplo, a idade dos participantes) explicam alguma da heterogeneidade dos efeitos de tratamento entre os estudos (Thompson & Higgins, 2002).

Modelos de efeitos fixos ou aleatórios

A meta-análise pode basear-se em qualquer um de dois modelos estatísticos – de efeitos fixos ou de efeitos aleatórios. O modelo de efeitos fixos assume que existe um verdadeiro efeito subjacente aos estudos em análise, que todas as diferenças nos dados são devidas a erro de amostragem ou ao acaso dentro de cada estudo e, ainda, que não há heterogeneidade entre os estudos. Ou seja, nos efeitos fixos é considerado que a variabilidade entre os estudos ocorreu apenas pelo acaso, ignorando-se a heterogeneidade entre eles (JBI, 2014a; Santos & Cunha, 2013).

O modelo de efeitos aleatórios assume que há uma população de estudos a partir da qual os estudos que são analisados na meta-análise foram incluídos na amostra de estudos e antecipa futuros estudos que venham a ser conduzidos, ou estudos já desenvolvidos que não foram identificados para a análise em causa. Assim, a inferência baseia-se no pressuposto de que os estudos utilizados na análise são uma amostra aleatória de uma população hipotética de estudos. Apesar dos efeitos dos estudos não serem considerados iguais, o seu conjunto apresenta uma distribuição de probabilidade normal (JBI, 2014a; Santos & Cunha, 2013).

Tufanaru et al. (2015, pp. 200-203) providenciam um fluxograma de decisão para a meta-análise referindo-se aos seguintes quatro critérios (Figura 1).

Como primeiro critério devemos considerar se existe a intenção de generalizar os resultados para além dos estudos incluídos. Neste caso, o modelo de efeitos aleatórios é a estatística adequada (assume-se que normalmente se quer generalizar as conclusões além dos estudos reais incluídos na meta-análise). Porém, havendo a intenção de considerar os resultados só para os estudos incluídos, sem generalizações, então o modelo de efeitos fixos é o modelo estatístico apropriado.

Um segundo critério a considerar refere-se diretamente ao número de estudos incluídos. O modelo de efeitos fixos é o adequado quando o número de estudos é pequeno pelo que deverá ser usado com um número inferior a cinco. Quando o número de estudos é suficiente em tamanho para suportar uma generalização para além dos estudos incluídos deve ser considerado o modelo de efeitos aleatórios.

O terceiro critério refere-se à heterogeneidade estatística. O modelo de efeitos fixos assume que todos os estudos incluídos estimam um único efeito verdadeiro subjacente. Se existe heterogeneidade estatística entre os tamanhos de efeito, então o modelo de efeitos fixos não é apropriado. Deve-se optar pelo modelo de efeitos aleatórios quando não pode ser assumido que a verdadeira homogeneidade existe.

O quarto critério relaciona-se com a probabilidade de um tamanho do efeito comum. Nos modelos de efeitos fixos, assume-se que há um efeito comum entre os estudos, e nos modelos de efeitos aleatórios assume-se que cada estudo estima um efeito subjacente

diferente e que os efeitos têm uma distribuição, geralmente, normal. O modelo de efeitos fixos só deve ser usado se for razoável supor que todos os estudos compartilham um efeito comum. Se nos estudos há heterogeneidade clínica e metodológica, não é razoável supor que partilham um efeito comum, sendo indicado o uso do modelo de efeitos aleatórios.

Outro critério refere-se à heterogeneidade dos tamanhos das amostras de estudos incluídos. O modelo de efeitos fixos é preferível quando um estudo inclui um elevado tamanho de amostra, já não quando há vários estudos com amostras pequenas (Tufanaru et al., 2015).

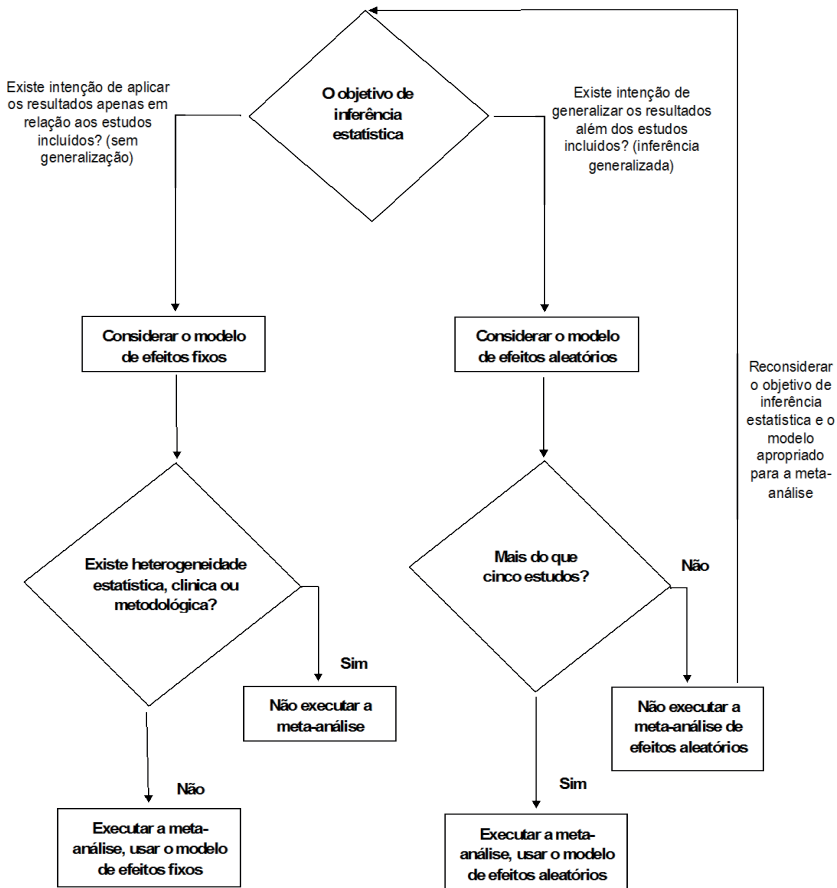


Figura 1. Fluxograma de decisão para a seleção do modelo estatístico de meta-análise. Adaptado de “Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness” por Tufanaru, Munn, Stephenson, e Aromataris, 2015, *International Journal of Evidence-Based Healthcare*, 13(3), p. 203.

Estatísticas na meta-análise

Há uma série de técnicas de meta-análise disponíveis para calcular um efeito global quando se combinam os resultados de diferentes estudos. A seleção de uma técnica² específica é orientada por três aspetos: o tipo de estudo, a natureza dos dados extraídos e os pressupostos subjacentes à meta-análise para dados contínuos e dicotómicos (JBI, 2014a).

Para dados contínuos sob um modelo de efeitos fixos existem três opções para a meta-análise:

- *Weighted mean difference* (WMD),
- *Hedge's adjusted standardized mean difference* (SMD),
- *Cohen's SMD*.

Para os dados contínuos sob um modelo de efeitos aleatórios, existem duas opções:

- *DerSimonian and Laird SMD*,
- *DerSimonian and Laird WMD*.

Para os dados dicotómicos sob um modelo de efeitos fixos, existem três opções:

- *Mantel-Haenszel Relative Risk* (RR),
- *Mantel-Haenszel Odds Ratio* (OR),
- *Peto OR*.

Para dados dicotómicos utilizando um modelo de efeitos aleatórios, há duas opções:

- *DerSimonian and Laird OR*,
- *DerSimonian and Laird RR*.

Dados contínuos

Quando estamos perante dados contínuos, a WMD é usada na meta-análise se em todos os estudos incluídos se mediu o mesmo resultado com o mesmo instrumento

2. Dada a multiplicidade de técnicas e combinação de termos, optámos por usar no texto o nome das estatísticas em inglês. No entanto, consideramos a seguinte tradução: *Weighted mean difference* (WMD): Diferença de médias ponderada; *Standardized mean difference* (SMD): Diferença de médias estandardizada; *Relative Risk* (RR): Risco relativo.

de medição. Não é calculada a diferença entre os valores iniciais e finais, mas sim a diferença global nas médias entre os grupos de tratamento em todos os estudos (equivalente ao tamanho do efeito) ponderado com o peso de cada estudo, peso este que está relacionado com os desvios padrão de cada estudo. Para o cálculo da WMD é utilizado o método inverso da variância para os modelos de efeitos fixos e o método de DerSimonian e Laird para modelos de efeitos aleatórios (JBI, 2014a).

Por sua vez, a SMD é usada na meta-análise de dados contínuos se nos estudos incluídos se mediu o mesmo resultado com diferentes instrumentos de medição (por exemplo diferentes escalas de medida para a dor). Nesta situação, os resultados são expressos em unidades de desvio padrão, sendo que os autores devem converter os resultados em unidades naturais/clínicas para facilitar a sua interpretação. Os resultados devem ser, idealmente, relatados quer em unidades de desvio padrão quer em unidades naturais. Em relação à opção pelo Cohen's ou Hedge's SMD estes produzem resultados semelhantes, mas o Hedge's é preferível porque inclui um ajuste para corrigir o viés de amostras pequenas (JBI 2014a; Tufanaru et al., 2015).

Dados dicotómicos

Quando o resultado de estudos incluídos é dicotómico, podem gerar-se dois tamanhos de efeito gerais com o cálculo do OR e do RR mas o OR é o método preferível para este tipo de dados. Ambos, OR e RR, podem ser combinados usando métodos Mantel-Haenszel. O Mantel-Haenszel é o método meta-analítico padrão para dados dicotómicos quando se opta por um modelo de efeitos fixos. O método de Mantel-Haenszel é geralmente preferido na meta-análise, comparativamente aos outros métodos, entre os quais o inverso da variância, porque é mais robusto quando os dados são escassos (em termos de baixas taxas de eventos e/ou o número de ensaios pequenos). Quando os dados são escassos, quer em termos de baixas taxas de eventos ou reduzido tamanho do estudo, as estimativas dos erros padrão e as estimativas de efeito que são utilizados nos métodos do inverso da variância podem ser pobres (JBI, 2014a; Tufanaru et al., 2015).

Quando se opta por um modelo de efeitos fixos, o Peto OR é um método alternativo para a meta-análise. Este método utiliza uma aproximação que pode ser imprecisa se os efeitos do tratamento são muito grandes e se os tamanhos de amostra entre os grupos de tratamento e controlo são desequilibrados. Mas este método é apropriado quando as taxas de eventos são muito baixas (OR perto de um) e os tamanhos de efeito não são muito grandes. Como estes critérios nem sempre são cumpridos, o método de Peto não é recomendado como uma abordagem padrão para meta-análise (Higgins & Green, 2011; JBI, 2014a).

Os métodos DerSimonian e Laird são usados na meta-análise de OR e RR usando um modelo de efeitos aleatórios. Uma variação do método do inverso da variância incorpora o pressuposto de que os diferentes estudos estimam diferentes, mas relacionados, efeitos da intervenção, produzindo uma meta-análise de efeitos aleatórios. Embora os tamanhos de efeito e estatística de heterogeneidade sejam calculados como para o modelo de efeitos fixos, os pesos dos estudos e os tamanhos de efeito gerais são calculados, neste modelo de efeitos aleatórios de DerSimonian e Laird, de forma ligeiramente diferente do que nos modelos fixos (JBI, 2014a).

Significância estatística, prática e clínica

Nas revisões sistemáticas é habitual que os autores se concentrem na significância estatística dos resultados. Tufanaru et al. (2015) recomendam que, para além da significância estatística (interpretada pelos investigadores), se deva dar atenção à significância prática e clínica. A significância prática (interpretada pelos investigadores e pelos clínicos) refere-se à magnitude e direção do tamanho do efeito, sendo, por exemplo, reportada em valores de OR. A significância clínica (interpretada pelos utilizadores) é considerada como a magnitude e direção do tamanho do efeito julgado pelos utilizadores, indicando um impacto positivo real na sua vida, por exemplo, através de uma melhoria real, de acordo com determinado critério de avaliação.

O gráfico de meta-análise

O método que mais se utiliza para apresentar os resultados da meta-análise é o gráfico de floresta (*forest plot*) que permite sumariar numa única figura todas as informações sobre o efeito/precisão do tratamento e a contribuição de cada estudo para a análise. Nestes gráficos é possível apresentar dados individuais dos estudos incluídos, resultados dos testes de heterogeneidade, resultados da meta-análise cuja opção deve obedecer à tipologia dos dados, contínuos ou dicotómicos, tipo de modelo de efeitos, fixos ou aleatórios.

Para cada estudo é apresentado um *bloco* correspondente à estimativa pontual de efeito da intervenção com uma linha horizontal que se estende de cada lado do bloco. A área do bloco indica o peso atribuído a esse estudo na meta-análise, enquanto a linha horizontal representa o IC (normalmente com um nível de 95% de confiança). A área do bloco e o intervalo de confiança dão informações semelhantes, mas ambos têm diferentes contribuições para o gráfico. O tamanho dos blocos evidencia os estudos com maiores pesos e usualmente com ICs mais estreitos (Higgins & Green, 2011).

Como se pode verificar na figura 2, o gráfico é ainda dividido por uma linha vertical, chamada linha de não efeito, que marca o efeito nulo. É possível observar dados dos participantes na totalidade e em cada grupo de tratamento, bem como o tipo de tratamento em causa. É possível observar, ainda, as estatísticas de heterogeneidade, o peso de cada estudo em análise, a estimativa pontual e intervalar, a estatística do efeito global ou para subgrupos e informação relativa ao modelo de efeitos utilizado.

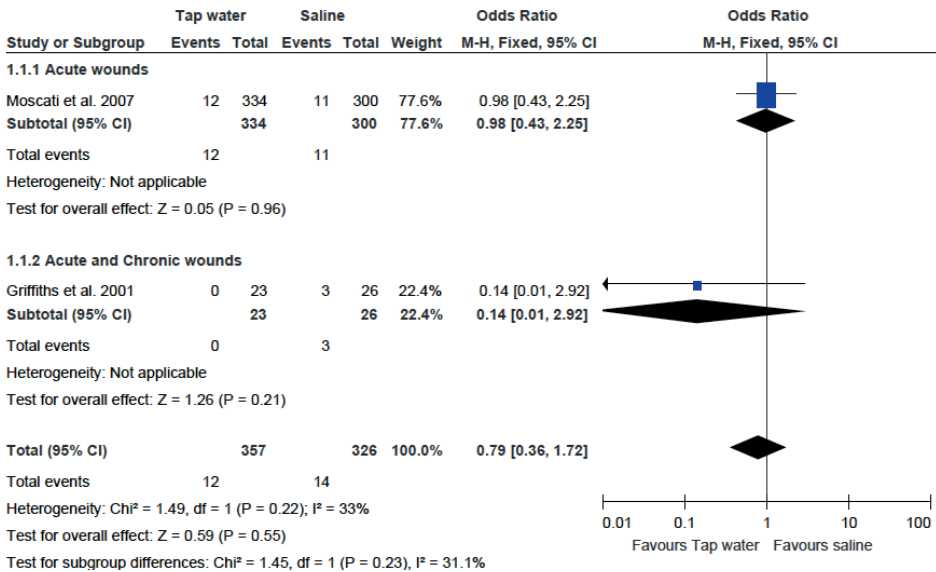


Figura 2. Gráfico de floresta comparando a eficácia da água da torneira versus a solução salina estéril sobre as taxas de infeção de feridas agudas e crónicas. Adaptado de “The effectiveness of cleansing solutions for wound treatment: A systematic review” por P. Queirós, E. Santos, J. Apóstolo, D. Cardoso, M. Cunha, e M. Rodrigues, 2014, *JBI Database of Systematic Reviews and Implementation Reports*, 12(10), p. 133.

O estudo representado na figura 2 (Queirós et al., 2014) apresenta fragilidades relacionadas com os pressupostos para a realização da meta-análise. Não obstante, e discutindo estas fragilidades, optámos pela sua realização.

Assim, apenas dois dos três estudos incluídos na RSL, com um total de 683 utentes, eram elegíveis para meta-análise. Ambos os estudos avaliaram a eficácia da água da torneira

versus a solução salina estéril e compararam as taxas de infecção da ferida. Os grupos de intervenção e de controlo foram comparados na linha de base e ambos os estudos eram semelhantes, excetuando-se a etiologia das feridas. Devido a essa variação, foi realizada uma meta-análise de subgrupos e o teste de diferenças de subgrupo mostrou uma heterogeneidade baixa (heterogeneidade Chi quadrado = 1,45; $p = 0,23$; $I^2 = 31,1\%$), cumprindo este pressuposto.

Para as feridas agudas, o risco (OR) de desenvolver uma infecção, quando a limpeza é feita com água da torneira em comparação com solução salina foi de 0,98 (IC 95%: 0,43- 2,25). A água da torneira foi mais eficaz do que a solução salina para reduzir a taxa de infecção em adultos com feridas agudas e crónicas (OR=0,14; IC 95%: 0,01-2,92).

A análise global estimou que não há diferenças estatisticamente significativas ($z = 0,59$; $p = 0,55$) entre a limpeza com água da torneira e com solução salina estéril sobre as taxas de infecção de ferida em feridas agudas e crónicas. No entanto, ainda podemos destacar que houve um efeito benéfico sobre o grupo de água da torneira na prevenção das taxas de infecção que é apoiado pelos resultados meta-analíticos (OR = 0,79; 95% IC: 0,36,-1,72). Deve também notar-se que dos dois estudos incluídos na meta-análise, o de Moscati et al. tem um peso mais elevado (77,6%) do que de Griffiths et al. (22,4%).

3.2 RSL de evidência de significado

O objetivo global das revisões que incluem estudos qualitativos é, sobretudo, o de informar sobre o significado que uma intervenção, procedimento, processo, estados ou condições de saúde têm para as pessoas num determinado contexto. Por esta razão a consideramos como RSL de evidência de significado.

A revisão de estudos de evidência de significado segue a mnemónica PICo: População, fenómeno de Interesse e Contexto (JBI, 2014a).

Título

A mnemónica PICo deve ser usada para construir um título claro de uma revisão sistemática de evidência de significado traduzindo a população, o fenómeno de interesse e o contexto, bem como o sentido compreensivo do significado, da experiência, entre outros (JBI, 2014a).

Exemplo: As experiências de conforto relacionadas com intervenções não farmacológicas de enfermagem de utentes com cancro.

Questão de revisão

O título anterior foi adaptado de uma revisão sistemática abrangente da literatura em desenvolvimento a partir de protocolo já publicado - *The use of non-pharmacological nursing interventions on the comfort of cancer patients: a comprehensive systematic review protocol* (Apóstolo et al., 2013). Esta revisão tem por objetivo identificar e sintetizar as melhores evidências disponíveis sobre a eficácia de intervenções não farmacológicas de enfermagem em doentes com cancro e as experiências de utentes relativamente às referidas intervenções. É abrangente por ter duas questões, uma quantitativa e uma qualitativa. A primeira questão pretende responder a aspetos de eficácia das intervenções; a segunda questão pretende responder a aspetos de significado destas intervenções para os utentes. Voltaremos a este exemplo no ponto sobre as revisões com métodos mistos (MM). Os exemplos que se seguem são baseados naquela referência.

População

A referência específica às características da população deve assentar numa justificação clara, científica e não num raciocínio não substancial, pessoal e sem fundamento. Não implica os aspetos da população que são pertinentes das revisões quantitativas, tais como métodos de amostragem, dimensão da amostra ou homogeneidade. Nestas revisões faz sentido referir-se na seleção da amostra que foram consideradas características dos participantes que asseguram que eles representam o fenómeno em análise. A exposição a uma doença ou intervenção, a interação com os profissionais de saúde, o relacionamento dos participantes com as experiências vividas são exemplos dos tipos de características da população que podem ser consideradas nestas revisões.

Exemplo: Uteses com cancro com 18 anos ou mais, com qualquer tipo e em qualquer estágio, a receber ou não tratamento farmacológico.

Fenómeno de interesse

O nível de detalhe atribuído aos fenómenos pode variar com a natureza ou complexidade do tema. Pode ser clarificada, expandida ou revista conforme o protocolo se vai desenvolvendo.

Exemplo: Experiências de conforto por utentes submetidos a intervenções não farmacológicas de enfermagem.

Contexto

Numa análise qualitativa, o contexto poderá variar de acordo com o objetivo da revisão e das perguntas específicas construídas para atingir o objetivo. O contexto pode incluir, mas não se limita a, considerações sobre fatores culturais (tais como localização geográfica, interesses de minorias ou de gênero), detalhes sobre o contexto de cuidados (como cuidados intensivos, cuidados primários de saúde ou a comunidade), bem como a sua relação com as experiências ou significados individuais ou grupais relatados nos estudos primários.

Exemplo: Utentes em quimioterapia em ambulatório ou hospitalizados.

Resultado

Como não há consenso internacional claro sobre o resultado nas RSLs qualitativas, não existe nenhuma exigência específica. Um resultado de interesse pode ser a descrição do fenómeno de interesse.

Exemplo: No caso concreto serão sintetizadas as experiências de conforto relativas às intervenções não farmacológicas de enfermagem.

Tipo de estudos

O desenho dos estudos a incluir deverá ser específico para dar resposta à questão de revisão. Nas revisões de significado são habitualmente incluídos estudos de natureza qualitativa de qualquer tipo.

Exemplo: Para esta revisão foram considerados quaisquer estudos interpretativos sobre as experiências de conforto de utentes com cancro submetidos a intervenções não farmacológicas de enfermagem. Podem incluir, mas não estão limitados a, estudos fenomenológicos, etnográficos ou de teoria fundamentada.

Síntese dos dados

Um só estudo qualitativo não deve ser considerado como constituinte de evidência substancial, tal como já referenciámos a propósito da investigação quantitativa. Por essa razão, os resultados de diversos estudos qualitativos, sempre que possível, devem ser combinados/ categorizados com base na semelhança de significado por um processo de agregação, tendo

por objetivo final gerar declarações para a prática. A síntese dos resultados qualitativos é geralmente referida como meta-síntese para distinguir da meta-análise quantitativa. Se a meta-síntese não for possível, os resultados serão apresentados em forma de narrativa.

A meta-síntese é um processo de combinar os resultados de estudos qualitativos individuais para criar declarações sumárias. É um processo interpretativo que exige transparência e requer especialização para identificar e extrair os resultados (*findings*) dos estudos incluídos na revisão, categorizar esses resultados e, finalmente, proceder à agregação das categorias num achado final (*synthesize finding*), que constitui uma declaração para a prática (JBI, 2014a; Pearson et al., 2005). Neste processo, há duas abordagens dominantes: a abordagem agregativa/integrativa e a abordagem interpretativa.

O modelo JBI de meta-síntese utiliza uma abordagem meta-agregativa para a síntese de evidências qualitativas (JBI, 2014a). A abordagem meta-agregadora do JBI é sensível à praticabilidade e usabilidade das conclusões do autor primário e não pretende voltar a interpretar (reinterpretar) esses resultados como outros métodos de síntese qualitativa advogam. Uma característica forte da abordagem do JBI é que procura ir além de um resultado de sugestões implícitas, a fim de produzir declarações ou diretivas práticas para orientar os profissionais e decisores (Hannes & Lockwood, 2011).

Como tal, a abordagem JBI contrasta com a abordagem da meta-etnografia, que tem um foco mais na interpretação do que na agregação. A utilidade de meta-etnografia reside na sua capacidade de gerar teoria que não é o que se pretende com a meta-síntese integrativa (Hannes & Lockwood, 2011; Pearson et al., 2005; Pearson, Field, & Jordan, 2007).

A meta-etnografia foi concebida, por Noblit e Hare (1988), como um método de síntese em que as interpretações podem ser construídas a partir de dois ou mais estudos etnográficos. Os autores consideram que os estudos etnográficos são atos interpretativos, e demonstram que, traduzindo as metáforas e os conceitos-chave dos diversos estudos etnográficos é possível desenvolver uma síntese interpretativa mais ampla.

A meta-etnografia envolve a identificação de estudos semelhantes na abordagem, método e foco através de um processo iterativo de análise. A abordagem agregativa envolve a *montagem* das conclusões dos estudos e a reunião dos resultados através da sua agregação com base na sua similaridade de significado (Hannes & Lockwood, 2011; JBI, 2014a; Pearson et al., 2007; Pearson et al., 2005).

A meta-síntese é a congregação textual de informações. É o processo de combinar as conclusões individuais de estudos qualitativos para criar uma síntese que autenticamente descreve o significado das categorias/temas. Tal como a meta-análise, a meta-síntese é parte integrante da revisão sistemática (Hannes & Lockwood, 2011; Lockwood et al., 2015; Pearson et al., 2005; Pearson et al., 2007).

3.3 RSL de prevalência e incidência

O objetivo global de revisões de dados de prevalência e incidência é informar sobre a frequência e distribuição de fatores específicos, estados de saúde ou condições numa população definida (JBI 2014b; Munn et al., 2015). Estas são importantes na descrição da distribuição geográfica de uma variável e variação entre os subgrupos (tais como género ou condição socioeconómica). Os resultados podem informar o planeamento dos cuidados de saúde e a alocação de recursos. A partilha de tais dados é necessária para monitorizar tendências relativas a uma doença e contribuir para a conceptualização de novos estudos etiológicos. As revisões sistemáticas são de particular relevância, porque os estudos individuais são habitualmente limitados pelo pequeno tamanho das amostras (JBI 2014b; Munn et al., 2015).

A revisão sistemática de estudos para responder a perguntas de dados de prevalência e incidência segue os mesmos princípios básicos das revisões sistemáticas de outros tipos de dados quantitativos. O objetivo destas revisões é descrever o problema de saúde (o quê), em que pessoas (quem), bem como o local (onde) e o período de tempo (quando) em que o problema de saúde ocorre ou ocorreu (JBI 2014b; Munn et al., 2015).

Título

A mnemónica CoCoPop deve ser usada para construir um título claro de uma revisão sistemática de evidência de prevalência e incidência, traduzindo a condição, o contexto e a população. O título deve dar uma indicação do tipo de dados que serão relatados (descritivo, analítico ou uma combinação de ambos), incluindo o indicador epidemiológico ou um termo que reflita a análise que irá ser usada para medir as variáveis de interesse. Geralmente, a medida da doença deve aparecer no título (prevalência ou incidência).

O título deve também especificar as características da população (ou seja, sexo, idade), bem como o local e hora da ocorrência, quando relevante (JBI 2014b; Munn et al., 2015).

Exemplo: *Prevalence of burnout in health professionals working in palliative care: a systematic review protocol* (Parola et al., 2016). Os exemplos que se seguem são baseados nesta referência.

Questão de revisão

A questão de revisão em estudos de evidência de prevalência e incidência segue a mnemónica CoCoPop - Condição, Contexto e População (JBI 2014b; Munn et al., 2015).

Exemplo: Qual é o efeito de trabalhar em cuidados paliativos no nível de *burnout* entre os profissionais de saúde, em comparação com os profissionais de saúde que trabalham noutros contextos?

População/Tipo de participantes

Características específicas da população podem incluir idade, sexo, etnia, *status* educacional, comportamento individual, fatores sociodemográficos, etc. (JBI 2014b; Munn et al., 2015).

Exemplo: Esta revisão considera estudos que incluam profissionais de saúde (médicos, enfermeiros, fisioterapeutas, assistentes sociais, terapeutas ocupacionais) a trabalhar em cuidados paliativos, que cuidam de utentes com 18 anos ou mais de idade. Estudos que incluam participantes que trabalham em outros contextos, que não os cuidados paliativos, serão excluídos.

Condição

Diz respeito à variável de interesse e pode referir-se a uma condição de saúde, doença, sintomas, evento ou fator que tem que ser claramente indicado e definido. Pode incluir o fornecimento de informações sobre a forma como a condição será medida, diagnosticada, ou confirmada, ou seja, por um exame de sangue, a observação de sintomas, uma lista de verificação validada, etc. (JBI 2014b; Munn et al., 2015).

Exemplo: Esta revisão irá considerar estudos que incluam a prevalência pontual de *burnout*, medido por qualquer escala de *burnout*, incluindo, mas não limitado a, *Maslach Burnout Inventory*, *Burnout Measure* e *Copenhagen Burnout Inventory*.

Contexto

Os fatores ambientais podem ter um impacto substancial sobre a prevalência ou incidência de uma condição sendo, por esta razão, importante a definição da situação ou contexto específico relevantes para a questão de revisão (JBI 2014b; Munn et al., 2015).

Exemplo: Esta revisão considera estudos realizados no contexto dos cuidados paliativos, mais especificamente, unidades de cuidados paliativos ou cuidados domiciliários.

Tipos de estudo

As RSL de prevalência e incidência são predominantemente derivadas de estudos observacionais. Desenhos de estudo transversais (para estudar a prevalência pontual

de condições comuns) ou estudos de coorte (com desenho prospetivo ou longitudinal e um *follow-up* de cada sujeito para estabelecer a incidência de uma doença ou a história natural de uma condição) podem fornecer informações sobre prevalência e incidência (JBI 2014b; Munn et al., 2015).

Exemplo: Esta revisão considera para a inclusão desenhos de estudos observacionais, incluindo estudos de coorte prospetivos e retrospectivos, caso-controlo e transversais.

Síntese dos dados

Os dados quantitativos de estudos de prevalência e incidência podem ser apresentados em forma de narrativa, incluindo tabelas e figuras, mas também combinados numa meta-análise estatística, caso se cumpram os pressupostos de utilização dos modelos e estatísticas já referidas anteriormente. Os gráficos de florestas, cumulativos ou parcelares (com ou sem uma estimativa combinada), podem ser particularmente úteis para apresentar como as estimativas de prevalência e incidência variam entre os estudos, populações e ao longo do tempo. Informação adicional sobre os procedimentos específicos de meta-análise nos estudos de prevalência e incidência, como a relativa à transformação de dados, é descrita por Munn et al. (2015).

3.4 RSL de associação (etiologia)

O objetivo global de revisões de dados de associação (etiologia), também referidos como correlacionais, é determinar se e até que ponto existe uma relação entre duas ou mais variáveis. Assim, a revisão deve delinear a exposição, doença, sintoma ou condição de saúde de interesse, da população ou grupos de risco, podendo incluir quaisquer fatores contextuais, como geográficos, temporais ou elementos culturais relevantes, ou o período de tempo, se relevante (Moola et al., 2015). A avaliação da associação incluirá as medidas de associação que podem indicar a magnitude e a direção da associação (Moola et al., 2015).

Sendo os estudos incluídos de associação entre variáveis, não se pode inferir diretamente a etiologia e efeito, uma vez que existem muitas variáveis que não são controladas. No entanto, é possível identificar fatores relacionados com o resultado/desfecho em saúde investigando correlações entre duas variáveis, por exemplo, entre uma variável e a evolução de uma doença/condição clínica. Neste sentido, as revisões de associação (etiologia) são importantes para o planeamento das políticas, dos processos e recursos na saúde, e da formação de pressupostos a respeito de risco ou fatores preventivos no desenvolvimento e progressão de uma doença ou condição clínica. Nestas revisões é necessário, então, determinar a validade da associação entre as variáveis em estudo (Moola et al., 2015).

Título

A mnemónica PEO deve ser usada para construir um título claro de uma revisão sistemática de associação (etiologia), traduzindo a População, a Exposição de interesse e o resultado/*Outcome*, bem como, a identificação da associação entre as variáveis ou fatores específicos, estados de saúde ou condições (Moola et al., 2015). Os exemplos que se seguem não estão publicados.

Exemplo: Influência do estado nutricional no risco de queda de idosos institucionalizados em estruturas residenciais: um protocolo de revisão sistemática.

Questão de revisão

A questão de revisão em estudos de evidência de associação (etiologia) segue a mnemónica PEO: População; Exposição de interesse (variável independente) resultado/*Outcome* (variável dependente; Moola et al., 2015).

Exemplo: Qual é o efeito do estado nutricional no risco de queda de idosos institucionalizados em estruturas residenciais para pessoas idosas?

População/Tipo de participantes

As características específicas da população deverão ser apropriadas para a revisão e explicadas no *background*, considerando as razões para a sua inclusão ou exclusão. Devem refletir o raciocínio clínico e científico relacionado com a homogeneidade das amostras nos estudos, tais como idade, sexo, variáveis de interesse ou localização/contexto, entre outras (Moola et al., 2015).

Exemplo: Esta revisão considera estudos que incluam idosos de ambos os sexos com mais de 65 anos a viverem em estruturas residenciais para idosos.

Exposição de interesse (variável independente)

A exposição de interesse (variável independente) pode referir-se a fatores associados com uma condição de saúde, doença, evento ou fator numa determinada população ou grupo a ela exposta que tem de ser claramente indicada e definida. Pode incluir o fornecimento de informações sobre a forma como a exposição de interesse terá sido avaliada (medição primária e/ou secundária). Tal como, na população os critérios de inclusão relacionados com a(s) variável(is) irão determinar o âmbito e algum grau de homogeneidade nos estudos a incluir (Moola et al., 2015).

Exemplo: Esta revisão irá considerar estudos que incluam a avaliação do estado nutricional de idosos, medido por qualquer escala ou critério, tais como, mas não limitado a, Mini Avaliação Nutricional, *Malnutrition Universal Screening Tool*, dobra cutânea tricipital, medida de peso, índice de massa corporal, força de prensão palmar, dobra cutânea subescapular ou circunferência da cintura.

Resultado/*Outcome* (variável dependente)

Deve considerar-se os resultados importantes de interesse, quando e como o resultado pode ser medido e, além disso, determinar se a avaliação deve examinar os desfechos secundários ou variáveis mediadoras (Moola et al., 2015).

Exemplo: Esta revisão considera estudos que incluíam os seguintes resultados: risco de queda medido por qualquer escala, como o Índice de Tinetti, Escala de Morse ou Escala St Thomas's *Risk Assessment Tool in Falling Elderly Inpatients*, entre outras. Serão consideradas variáveis mediadoras como a idade, visão, uso de medicamentos, doenças associadas, flexibilidade, força e equilíbrio, entre outras.

Tipos de estudos

As revisões de associação (etiologia) são predominantemente derivadas de estudos observacionais. Estes incluem estudos retrospectivos, prospectivos, transversais, longitudinais, caso-controle e estudos de coorte. Os estudos randomizados controlados podem também informar sobre o risco associado a uma intervenção e podem ser incluídos neste tipo de revisões (Moola et al., 2015).

Síntese dos dados

Tal como acontece com todas as revisões sistemáticas, há várias abordagens para a apresentação dos resultados, incluindo a apresentação narrativa, com recurso a gráficos ou tabelas de resumo e a meta-análise. A meta-análise de estudos de associação que endereçam etiologia raramente pode ser realizada pelas diferenças nos fatores controlados nas análises multivariadas, e, também, por causa da falta de detalhes adequados nos estudos publicados. A meta-análise é apropriada somente quando os estudos são suficientemente homogêneos do ponto de vista clínico e metodológico, pelo que o critério estatístico não deverá ser o orientador *major* da decisão, tal como já referimos anteriormente.

O modelo de efeitos aleatórios é mais ajustado porque *compensa* a heterogeneidade aparente em estudos de observação, conseguindo estimar a variação inexplicável na análise. Quando não há heterogeneidade presente, os resultados dos modelos de efeitos de fixos e aleatórios serão semelhantes (Moola et al., 2015). Outros aspetos relativos às estatísticas a utilizar foram referidos anteriormente.

3.5 RSL de precisão de testes de diagnóstico

As revisões sistemáticas de precisão de testes de diagnóstico fornecem um resumo do desempenho dum teste com base em todas as evidências disponíveis e contabilizam a variação dos resultados entre os estudos (Campbell et al., 2015; JBI, 2015b).

Neste ponto não daremos exemplos concretos desta metodologia. Estes serão apresentados seguidamente no ponto respeitante à revisão de revisões sobre precisão de testes de diagnóstico. As considerações que se seguem sobre esta tipologia de revisões podem ser consultadas em documentos de autoria do JBI (2015b) e Campbell e colaboradores (2015).

Título

A mnemónica PIRD deve ser usada para construir um título claro de uma revisão sistemática de precisão de testes de diagnóstico, traduzindo a População, o teste diagnóstico de Interesse (*Index test*), o teste de Referência (*Reference test*) e o Diagnóstico de interesse. A precisão do teste de interesse deve ser comparada com o padrão ouro.

Questão de revisão

A questão de revisão de estudos sobre precisão de testes de diagnóstico segue a mnemónica PIRD: População, teste diagnóstico de Interesse, teste de Referência e Diagnóstico de interesse.

População/Tipo de participantes

O tipo de participantes deve ser apropriado para os objetivos de revisão e refletir que serão submetidos a teste de diagnóstico na prática clínica. Devem ser ainda consideradas as características da população que podem ser relevantes, incluindo o estágio da doença, sintomas, idade, sexo, etnia, nível educacional, etc.

Teste diagnóstico de Interesse

O teste diagnóstico de interesse é o teste cujas propriedades estão a ser alvo de investigação na revisão.

Teste de referência

O teste de referência é o *padrão de ouro* com o qual os resultados do teste diagnóstico de interesse serão comparados. Deve ser o(s) melhor(es) teste(s) disponível(eis) para o diagnóstico da condição de interesse.

Diagnóstico de interesse

Este item refere-se ao diagnóstico alvo da RSL podendo ser uma doença, lesão, deficiência ou qualquer outra condição de saúde.

Tipos de estudos

É feita uma descrição do tipo de estudos que serão considerados para inclusão na revisão. Nos estudos de diagnóstico são geralmente usados desenhos de estudos transversais ou de caso-controlo.

Síntese dos dados

As estimativas da precisão do teste variam, frequentemente, entre os estudos, muitas vezes devido a diferenças na forma como o teste foi positivamente definido, no desenho do estudo, nas características da amostra e no posicionamento do teste no procedimento diagnóstico (Campbell, et al. 2015; JBI, 2015b). Os estudos de precisão de diagnóstico têm características únicas de desenho e relatam um par de estatísticas de sensibilidade e de especificidade ao invés de uma única estatística. Consequentemente, as revisões com esta tipologia requerem diferentes métodos estatísticos de meta-análise e, também, diferentes abordagens para a síntese narrativa. A meta-análise é mais complicada do que a maioria das outras formas de RSL, principalmente, devido à natureza emparelhada das principais medidas de resultado, isso é, sensibilidade e especificidade.

Cada estudo primário contribui para um ponto singular definido pela sua sensibilidade e especificidade para um determinado limiar diagnóstico. Visto que a sensibilidade e

a especificidade dependem da seleção de um limiar de decisão, a análise *Receiver Operating Characteristic* (ROC) é usada para representar graficamente a sensibilidade (eixo dos y) contra 1 - (menos) a especificidade (eixo dos x) e as respetivas mudanças de valor limiar (Campbell et al., 2015; JBI, 2015b).

Nas revisões sistemáticas de estudos de precisão de diagnóstico é frequente encontrar heterogeneidade entre os estudos a qual pode ser devida a diferenças nas populações do estudo, procedimentos seguidos para a realização do teste (de diagnóstico ou de referência) e condições ou contexto da sua aplicação. Como tal, a presença de heterogeneidade entre os estudos deve ser cuidadosamente investigada.

O teste do qui-quadrado ou teste exato de Fisher pode ser usado para avaliações mais objetivas de heterogeneidade, porém, o seu poder é considerado baixo. O teste I^2 não é rotineiramente utilizado na RSL de precisão do teste diagnóstico, uma vez que não é tida em conta a influência de diferentes limiares de decisão. A análise de subgrupos pode ser utilizada para investigar potenciais fontes de heterogeneidade. Contudo, quando a extensão e causa da heterogeneidade não podem ser explicadas, os revisores devem abster-se de desenvolver meta-análise e, em vez disso, realizar uma síntese narrativa (Campbell et al., 2015; JBI, 2015b).

3.6 RSL de métodos mistos (*Mixed Methods*)

A síntese mista de investigação, no que respeita a revisões sistemáticas, têm por objetivo a integração de estudos quantitativos e qualitativos em domínios comuns de investigação (Sandelowski et al., 2006) visando agregar, integrar, ou reunir de outra forma os seus resultados através do uso de métodos qualitativos e/ou quantitativos (Sandelowski, Voils, & Barroso, 2007)

Os interesses de utilizadores das evidências são muito variados e vão desde os efeitos de uma intervenção específica até à experiência subjetiva dos utentes ou profissionais de saúde. Até mesmo nas investigações que ostensivamente estudam as mesmas variáveis ou eventos alvo, experiências ou fenómenos em grupos aparentemente semelhantes, ou que empregam deliberadamente as mesmas abordagens metodológicas, poderão existir diferenças suficientes que justificavam o desenvolvimento de métodos que permitam a comparação e combinação dos seus resultados (Sandelowski et al., 2006).

Como temos vindo a exemplificar ao longo deste texto, dependendo do tipo de evidência que se pretende sintetizar, a metodologia tem que ser ajustada às diferentes questões de revisão. No entanto, pode haver necessidade de responder a questões que necessitam de metodologia mista. Num determinado contexto de saúde podem colocar-se questões

comuns tais como: Qual é o problema?; Como funciona?; É mais importante para algumas pessoas do que para outras?; É aceitável este procedimento?; Funciona?; O que significa para o utente?; Qual será o melhor processo para obter apoio político?; O que funciona melhor numa determinada população?; Qual a relação custo benefício, custo eficácia, custo utilidade ou custo minimização?. Assim, por exemplo, mesmo que a evidência suporte a eficácia de um medicamento ou procedimento específico, os decisores ou utilizadores da evidência podem ainda querer saber se o uso desta intervenção é viável (tanto em termos práticos, como financeiros), se a intervenção é culturalmente e eticamente adequada num dado contexto e o que significa para os utentes ou consumidores.

Portanto, na área da saúde, determinados focos ou questões complexas ou com diferentes dimensões podem necessitar de métodos mistos para produzir uma resposta condizente. As revisões que utilizam MM poderão ser utilizadas para sintetizar evidência incluindo estudos que, apesar de terem sido desenvolvidos com diferentes desenhos de investigação, são centrados no mesmo tópico. Os diferentes resultados ou achados são depois combinados/agregados. Ao incluir diversas formas de evidências de diferentes tipos de investigação, as revisões de MM pretendem maximizar os resultados e a capacidade dessas descobertas para informar a política e a prática (JBI, 2014c; Pearson et al., 2014; Pearson et al., 2015).

A síntese meta-agregadora adotada pelo JBI assenta na abordagem Bayesiana em que os dados quantitativos são convertidos em qualitativos, como proposto por Crandell, Voils, Chang, e Sandelowski (2011).

Para manter o rigor na tradução de resultados quantitativos em descrições textuais qualitativas é importante que se preste atenção à minimização do possível impacto deste processo. Tal como na síntese de evidência qualitativa nas RSLs com método único, os revisores deverão considerar a minimização do impacto da síntese, *colocando entre parênteses* os seus próprios interesses, experiência pessoal ou crenças e aspetos culturais que possam influenciar a (re)interpretação dos resultados. Na síntese de MM, a extração das categorias dos estudos incluídos é feita a um nível mais elevado, desaconselhando-se a reinterpretação dos achados primários (JBI, 2014c; Pearson et al., 2014; Pearson et al., 2015).

Neste processo, deve também assegurar-se a ligação ao contexto, procedimento que Sandelowski, Leeman, Knafl, e Crandell (2013) referem como *texto-em-contexto*. Para preservar o texto-em-contexto dos resultados obtidos, os referidos autores descrevem o método pelo qual as conclusões são transformadas em declarações para a prática, apresentando as seguintes regras de ouro para a ancoragem do texto: informações relevantes sobre amostra, fonte de informação, tempo, ponto de referência comparativa entre grupos e entre temas, magnitude, importância (frequência, significado, percentagem de participantes associada a um achado) e conceptualizações específicas do estudo (fenómenos alvo, tal como operacionalizado no estudo).

Como referimos, para que as sínteses provenientes dos diferentes desenhos de estudos possam ser combinadas, os dados quantitativos são transformados em informação qualitativa (qualificação de dados quantitativos), ou seja, em descrições textuais temáticas, sendo colocados lado a lado com os resultados sintetizados gerados a partir da análise qualitativa. Finalmente, estas descrições textuais e achados sintetizados são agregados ou *casados* para gerar uma síntese a partir de MM (JBI, 2014c; Pearson et al., 2014; Pearson et al., 2015).

Título

Por definição, uma revisão de MM incorpora pelo menos dois métodos únicos (por exemplo, uma síntese com foco em ensaios clínicos randomizados para medir os efeitos de uma intervenção e uma síntese com foco em estudos qualitativos que exploram as experiências dos utentes). O seu desenvolvimento aborda dois ou mais critérios que estão alinhados com os respetivos métodos únicos. A abordagem *segregadora* adotada pelo JBI mantém uma clara distinção entre os dados quantitativos e qualitativos e requer sínteses individuais a serem realizadas antes da combinação final (JBI, 2014c; Pearson et al., 2014; Pearson et al., 2015; Sandelowski et al., 2006).

A mnemónica a ser usada para construir um título claro de uma revisão sistemática com MM depende do tipo de estudos que se pretendem combinar. Retomamos, aqui, o exemplo utilizado nas revisões de significado (Apóstolo et al., 2013), adaptando-o como uma revisão de métodos mistos. Os exemplos que se seguem são adaptados a partir desta referência.

Exemplo: O uso de intervenções não farmacológicas de enfermagem no conforto dos utentes com cancro: uma revisão sistemática de métodos mistos.

Questão de revisão

A mnemónica a ser usada para construir as questões de revisão sistemática de MM depende do tipo de estudos que se quer combinar e requer a articulação clara das questões.

Exemplo: Esta revisão tem por objetivo identificar e sintetizar as melhores evidências disponíveis sobre a eficácia de intervenções não farmacológicas de enfermagem em doentes com cancro e as experiências dos utentes face às referidas intervenções.

Mais especificamente, a revisão incide sobre as seguintes questões:

- São as intervenções não farmacológicas de enfermagem eficazes para aumentar o conforto em utentes com cancro?

- Quais são as experiências dos utentes com cancro relativamente a intervenções não farmacológicas de enfermagem? Como experienciam os utentes com cancro as intervenções não farmacológicas de enfermagem?

População/Tipo de participantes

Exemplo: As componentes, quantitativa e qualitativa, desta revisão consideram estudos que incluíram utentes com cancro, com idade de 18 ou mais anos, com qualquer tipo e em qualquer estágio do cancro, recebendo ou não tratamento farmacológico.

Tipos de intervenção/ Fenómenos de interesse/Contexto

Exemplo: Esta revisão considera estudos que utilizaram intervenções não farmacológicas de enfermagem para adultos com cancro (componente quantitativa) e/ou que se concentraram nas experiências de conforto de utentes submetidos a intervenções não farmacológicas de enfermagem (componente qualitativa) em contexto ambulatorio e hospitalar.

Tipo de estudos

Nestas revisões podem ser consideradas diversas tipologias de estudos quantitativos e qualitativos que possam dar resposta à questão de revisão.

Exemplo: Esta revisão considera qualquer desenho experimental, incluindo ensaios clínicos randomizados ou não randomizados ou outros estudos quase-experimentais, tais como, estudos antes-após (componente quantitativa). Considera, ainda, qualquer estudo qualitativo/interpretativo sobre as experiências de utentes com cancro sobre conforto quando submetidos a intervenções não farmacológicas de enfermagem (componente qualitativa).

Tipo de resultados

Exemplo: Esta revisão considera estudos que incluíram os resultados de conforto e outros resultados relacionados com este conceito, medidos por qualquer escala de conforto, dor, ansiedade, depressão, stresse ou fadiga (quantitativo), bem como as experiências de conforto relativas às intervenções não farmacológicas de enfermagem (qualitativo).

Síntese dos achados

Os resultados (*findings*), fruto de cada uma das sínteses de métodos únicos, são *montados/agregados*. Este processo envolve a sua configuração para gerar um conjunto de declarações que representam a agregação.

Os dados quantitativos são transformados em descrições textuais para poderem ser, de seguida, montados com os achados provenientes da agregação qualitativa. A *montagem* de todos os temas resultantes da síntese quantitativa e qualitativa vai produzir um conjunto de resultados sintetizados em forma de conclusões ou recomendações para a prática.

O método proposto pelo JBI para a síntese dos dados/achados (JBI, 2014c; Pearson et al., 2014; Pearson et al., 2015) é desenvolvido em cinco passos que apresentamos de seguida.

1. Identificar os resultados quantitativos. Deve ler-se a síntese quantitativa para identificar quaisquer resultados importantes que informam o tema de interesse. Deverão incluir-se todas as estatísticas (quando disponíveis) e identificar os aspetos contextuais que podem afetar a força da descrição textual resultante.

2. Converter os dados quantitativos em descrições textuais qualitativas semelhantes e colocá-los numa tabela de conversão.

3. Apresentar uma discussão aprofundada das descrições textuais quantitativas. Como ocorre nas sínteses qualitativas, cada descrição textual obtida através da análise dos resultados quantitativos convertidos necessita de ser discutida em contexto.

4. As afirmações quantitativas e qualitativas devem ser caracterizadas. Tal como na investigação qualitativa, em que se combinam categorias para formar resultados sintetizados, também aqui as descrições textuais resultantes da qualificação de resultados quantitativos são combinados/casados com os resultados sintetizados da meta-síntese qualitativa para formar uma ou mais declarações para a prática. Os resultados qualitativos podem confirmar ou refutar as conclusões quantitativas ou complementares.

5. Discutir aprofundadamente a síntese combinada final, incluindo quaisquer declarações contextuais que possam afetar a generalização ou a aplicabilidade dos resultados.

O resultado final, em vez de assumir uma abordagem prescritiva ou um somatório dos resultados, deve apontar para o fornecimento indicativo de instruções com base na evidência disponível.

3.7 RSL de revisões de revisões (*Umbrella*)

Uma revisão de revisões pode ter diversos nomes na literatura científica, como: revisões guarda-chuva (*umbrella review*), visões gerais de revisões, revisões de revisões, um resumo das revisões sistemáticas. Considerando o grande número de revisões sistemáticas disponíveis para informar a prática ou política na área da saúde, as revisões de revisões podem ser desenvolvidas com o objetivo de comparar e contrastar a evidência gerada pelas RSLs publicadas num determinado tópico de interesse, fornecendo uma análise global do corpo de informação disponível (JBI, 2014d).

A mnemónica deve fornecer uma quantidade significativa de informações sobre o foco, o alcance e a aplicabilidade da revisão de revisões. O título de uma revisão de revisões deve ser suficientemente conciso, mas também descritivo, para refletir as intervenções ou os fenómenos de interesse como um todo. A questão de revisão seguirá a mnemónica da tipologia das RSLs originais tal como já referimos anteriormente. Nas revisões de revisões devem ser, também, detalhadas as características importantes dos participantes, incluindo a idade e outros critérios de qualificação, que os tornam apropriados para os objetivos da revisão de revisões e que são condizentes com a pergunta de revisão. Deve ser coerente com o título e dirigir o desenvolvimento dos critérios de inclusão específicos claramente identificáveis (JBI, 2014d).

Sem mais detalhes, porque cada revisão de revisões deverá obedecer a especificidades relacionadas com a tipologia de RSLs a incluir, apresentamos exemplos de um protocolo de uma revisão de revisões sobre precisão de testes de diagnóstico relativamente à fragilidade na pessoa idosa (Apóstolo, Cooke et al., 2015).

Título

Exemplo: *Predicting risk and outcomes for frail older adults: a protocol for an umbrella review of available frailty screening tools*. No protocolo original foi seguida a mnemónica PICO contendo, contudo, os elementos da PIRD adaptados a este formato, que aqui apresentamos.

Questão de revisão

Exemplo: - Qual é a confiabilidade e a validade de instrumentos de rastreio existentes para avaliar a pré-fragilidade/fragilidade em idosos?

- Qual a sensibilidade e a especificidade destes instrumentos para prognosticar a evolução da fragilidade?

- Qual é a capacidade preditiva destes instrumentos no que respeita a risco de resultados adversos de saúde (incapacidade funcional, hospitalização, institucionalização, comorbidades e morte)?

População/Tipo de participantes

Exemplo: Esta revisão de revisões considera revisões sistemáticas que incluem idosos, com idades de 65 ou mais anos, de qualquer contexto, incluindo a rede de cuidados de saúde primários, hospitais, estruturas residenciais para idosos ou centros de dia.

Teste diagnóstico de interesse

Exemplo: Instrumentos de rastreio da pré-fragilidade e fragilidade em idosos. Serão analisados dados da qualidade do teste de interesse relativamente à sua confiabilidade (consistência interna e repetibilidade dos resultados), validade de critério e de constructo, sensibilidade e especificidade, e capacidade preditiva do risco de resultados adversos de saúde.

Teste de referência

Exemplo: Modelo fenotípico do *Cardiovascular Health Study* (CHS; Fried et al., 2001), modelo dos défices cumulativos do *Canadian Study of Health and Aging* (CSHA; Rockwood & Mitnitski, 2011) ou outro padrão ouro reconhecido, como o *Comprehensive Geriatric Assessment* (Rubenstein, Stuck, Siu, & Wieland, 1991).

Diagnóstico de interesse

Exemplo: Fragilidade ou pré-fragilidade considerando as diferentes definições: focadas em marcadores físicos de fragilidade (Fried et al., 2001); focadas nos défices cumulativos do domínio físico, cognitivo, de saúde mental e dos vários domínios funcionais (Rockwood & Mitnitski, 2011); que integram informação demográfica, médica, psicológica, social e funcional, tal como descrito por Pijpers, Ferreira, Stehouwer, e Nieuwenhuijzen Kruseman (2012).

Tipo de estudos

Nas revisões de revisões são habitualmente incluídas RSLs ou meta-análises.

Exemplo: Esta revisão de revisões considera revisões sistemáticas quantitativas e meta-análises para identificar evidência científica relevante relativamente a instrumentos de avaliação de fragilidade e pré-fragilidade no que respeita à sua confiabilidade, validade de critério, capacidade de detetar fragilidade e pré-fragilidade e prever desfechos de saúde adversos.

Sumário dos resultados

O objetivo da revisão de revisões é apresentar um resumo das sínteses de investigação existentes, relevantes para um determinado tópico ou pergunta, não uma síntese dos resultados destas publicações. Para este fim, os resultados de todos os estudos incluídos devem ser apresentados, permitindo uma visão geral e facilmente interpretável (JBI, 2014d).

É recomendada a apresentação tabular dos resultados quando se enumeram as estimativas de efeito gerais ou outros dados numéricos similares extraídos de revisões sistemáticas quantitativas. Devem, também, ser apresentados os dados referentes a: número de estudos que informam cada resultado (*outcome*), número de participantes, e heterogeneidade dos resultados das revisões incluídas.

Quando as revisões de revisões incluem resultados de RSLs qualitativas devem ser apresentados os achados globais ou sintetizados, em formato tabular e com a informação contextual relevante de cada achado sintetizado para garantir a sua interpretação.

Devem ser dadas indicações claras de quaisquer sobreposições de estudos primários incluídos nas RSL que foram incluídas na revisão de revisões (JBI, 2014d). Para além destes aspetos gerais, o sumário dos resultados numa revisão de revisões deverá obedecer às especificidades de cada tipologia de revisões incluídas. No exemplo em análise (Apóstolo, Cooke et al., 2015) foi considerada a possibilidade de recalcular as razões de verosimilhança positiva e negativa, OR e IC a 95%. Foi ainda considerada a apresentação de gráficos de floresta e ROC, e a avaliação estatística da heterogeneidade.

3.8 Revisão *scoping*

O conceito *scoping* não é de fácil tradução para a língua portuguesa. Deriva do latim (*scopus*: i) e do grego (*skopós*: óu, observador, espião, vigilante) e significa local bem determinado a que se aponta para atingir, alvo ou mira, objetivo que se pretende atingir, desígnio, fim, intuito, propósito, limite ou abrangência de uma operação (definição da palavra “escopo”, conforme Dicionário Priberam da Língua Portuguesa, recuperado de <http://www.priberam.pt/dlpo/escopo>). Assim, e porque não encontramos na língua portuguesa um termo que signifique o conceito, optamos por utilizar o termo original *scoping*.

Arksey e O'Malley (2005) consideram que é possível identificar pelo menos quatro razões comuns para desenvolver as revisões *scoping* (RS), a saber:

- Examinar a extensão, alcance e natureza da atividade de investigação. Este tipo de avaliação rápida pode não descrever os resultados da investigação em detalhe, mas é uma maneira útil para mapear áreas de estudo onde é difícil visualizar a gama de informação que pode estar disponível.
- Determinar a valia de realizar uma revisão sistemática. Um mapeamento preliminar da literatura pode gerar informação sobre a viabilidade e relevância da realização de uma RSL.
- Resumir e divulgar os resultados da investigação. Uma RS pode ajudar a descrever detalhadamente, bem como sumariar e divulgar os resultados da investigação numa determinada área às partes interessadas que não têm os recursos e capacidades para a desenvolver.
- Identificar as lacunas da investigação numa determinada área permitindo concluir sobre o estado global da atividade de investigação nessa área. No entanto, é importante notar que a identificação de lacunas na literatura através desta metodologia não vai necessariamente identificar as lacunas sobre a falta de qualidade da investigação, já que a avaliação da qualidade não faz parte deste tipo de revisões.

A justificação da necessidade do desenvolvimento é um dos mais complexos exercícios quando se desenvolve um protocolo de uma RS. É relativamente fácil providenciar no *background* a justificação da necessidade e utilidade para uma revisão sistemática de eficácia ou de outra metodologia, mas não assim tão fácil justificar a necessidade, utilidade e o valor de mapear uma determinada área do conhecimento. Este valor e necessidade podem ser justificados, por exemplo, afirmando as lacunas existentes na avaliação ou intervenção numa área ou foco, dispersão do conhecimento e explicando que a revisão vai conseguir mapear de uma forma específica para colmatar essas lacunas. Pode ainda ser reconhecido o valor da revisão para estabelecer as bases para uma revisão sistemática subsequente.

Outros tópicos que podem ajudar a justificar a necessidade de orientar a revisão são: Que evidência/conhecimento necessita ser examinada e mapeada? Quais os aspetos específicos sobre o estado da evidência em causa que impedem uma revisão sistemática quantitativa ou qualitativa? A evidência dispersa no universo do conhecimento (por exemplo, inclui uma mistura diversificada e heterogénea de intervenções, populações, abordagens, terminologia, etc.) impede o desenvolvimento direto de uma revisão sistemática quantitativa ou qualitativa? Existem questões importantes sobre a natureza dos elementos que necessitam de resposta prévia antes que uma questão objetiva de RSL possa ser enunciada?

A este propósito de justificar as RSs podem ser consultados alguns dos argumentos referidos em três protocolos já desenvolvidos (Coelho, Parola, Cardoso, Escobar, & Apóstolo, 2016; Lomba, Kroll, Apóstolo, Gameiro, & Apóstolo, 2016; Pedreiro, Loureiro, Ferreira, Cardoso, & Apóstolo, 2016).

Título

O título deve ser informativo e dar uma indicação clara do tema e da abrangência da revisão. A mnemónica PCC deve ser usada para construir um título claro de uma RS, traduzindo a População, o Conceito e o Contexto (JBI, 2015c).

Exemplo: The use of non-pharmacological interventions for the comfort of patients in palliative care: a scoping review protocol (Coelho et al., 2016). Os exemplos que se seguem são baseados nesta referência.

Questão de revisão

Numa RS a questão de revisão segue a mnemónica PCC: População, Conceito e Contexto. Esta é fundamental para organizar a seleção dos estudos, extração de dados, mapear a evidência e explicar o seu significado.

Exemplo: O objetivo desta RS é examinar e mapear as intervenções não farmacológicas implementadas e avaliadas para proporcionar conforto em cuidados paliativos.

Mais especificamente, a revisão incidirá sobre as seguintes questões:

- Que intervenções não farmacológicas foram implementadas e avaliadas para proporcionar conforto em utentes com doenças incuráveis e avançadas?
- Quais são as características (duração, dose e a frequência) dessas intervenções?
- Em que contextos (domicílio, unidade de cuidados continuados ou de cuidados paliativos) são as intervenções não farmacológicas implementadas e avaliadas?
- Em que populações (utentes com ou sem cancro) são as intervenções não farmacológicas implementadas e avaliadas?

População/Tipo de participantes

As características importantes dos participantes devem ser detalhadas, incluindo a idade e outros critérios de qualificação adequados aos objetivos/questão da RS. A

justificação da inclusão ou exclusão de participantes deve ser explicada. Podem também ser especificados fatores populacionais confundentes como morbilidades ou condições de estado de saúde (JBI, 2015c).

Exemplo: Esta revisão considera todos os estudos que se concentram em utentes com doença avançada incurável, com 18 ou mais anos, assistidos por equipas de cuidados paliativos.

Conceito

O conceito básico examinado pela RS deve ser claramente articulado para orientar o seu desígnio e amplitude. Pode incluir detalhes que dizem respeito a elementos também considerados numa revisão sistemática quantitativa ou qualitativa, tais como as intervenções e ou fenómenos de interesse (JBI, 2015c). No caso exemplificado sobre as intervenções não farmacológicas, implementadas e avaliadas para proporcionar conforto em utentes em cuidados paliativos (Coelho et al., 2016), foram referenciados os conceitos em análise incorporados nas questões de revisão. Entre eles: intervenções e as suas características, condições clínicas associadas, bem como os critérios relativos à avaliação das intervenções. Os resultados podem também ser uma componente de uma RS e devem estar estreitamente ligados ao objetivo e finalidade da revisão.

Exemplo: São parte integrante do conceito as diversas intervenções não farmacológicas incluindo, mas não limitado a, imaginação guiada, relaxamento, toque terapêutico e massagem. São ainda considerados o conceito de conforto e conceitos associados, como dor, sofrimento, ansiedade, bem-estar, stresse ou fadiga, e os instrumentos ou critérios para a sua medição/avaliação.

Contexto

O elemento de contexto de uma RS pode variar de acordo com o objetivo e as questões de revisão. Deve ser claramente definido e pode incluir, mas não está limitado a fatores culturais, geográficos ou de género ou ainda, contextos da prática clínica, hospitalar, comunitária ou educacional (JBI, 2015c).

Exemplo: Esta revisão considera todas as intervenções não farmacológicas realizadas e avaliadas no contexto de cuidados paliativos, incluindo o domicílio ou unidades de cuidados continuados ou paliativos.

Tipos de estudos/fontes

Nas RSs a fonte de informação pode incluir qualquer literatura existente, entre outros, estudos primários de investigação, revisões sistemáticas, textos, relatórios ou orientações, desde que sejam ajustadas à informação que se tenta mapear (JBI, 2015c).

Exemplo: Esta revisão considera estudos quantitativos, qualitativos e revisões sistemáticas. Os desenhos quantitativos incluem quaisquer desenhos de estudos experimentais (incluindo ensaios clínicos randomizados controlados, ensaios clínicos controlados não randomizados ou outros estudos quase-experimentais, tais como estudos antes-após) e estudos observacionais (estudos descritivos, estudos de coorte, estudos transversais, estudos de caso e de séries de casos). Os desenhos qualitativos incluem quaisquer estudos que se concentrem em dados qualitativos, tais como, mas não limitados, a fenomenologia, teoria fundamentada e desenhos de etnografia, entre outros.

Síntese dos dados

Este processo oferece ao leitor um resumo lógico e descritivo dos resultados que se alinham com o objetivo e a pergunta da revisão. Os resultados podem ser apresentados como um mapa dos dados extraídos dos artigos, sendo incluídos em diagrama(s) ou tabela(s) e em formato descritivo. Os elementos dos critérios de inclusão podem ser úteis para orientar a forma como a informação é mapeada e organizada de acordo com os diferentes elementos constituintes da população, do conceito ou do contexto (JBI, 2015c).

No exemplo em análise (Coelho et al., 2016) foram constituídas quatro tabelas, uma para cada questão, em que a informação é descrita em relação às intervenções, sintomas, contextos e tipo de doença.

Como foi descrito as diferentes questões orientam para diferentes tipologias de revisão e a adoção de metodologias também diversas. A ideia que passa na literatura quando é utilizada a clássica mnemónica PICO para dar resposta a todas as questões de revisão não parece ser a ideal. Cada questão deverá ser orientada por uma questão e metodologia diferente.

CAPÍTULO IV

RESULTADOS DE AVALIAÇÃO CRÍTICA E DISCUSSÃO

Neste documento temos vindo a aprofundar, sobretudo, a síntese no contexto do ciclo da translação da ciência. Assim, e após termos tecido considerações sobre a minimização do risco de viés nos estudos quantitativos e qualitativos; termos apresentado as diferentes tipologias de RSLs, as suas especificidades no que respeita ao título, estrutura da questão de revisão relativamente a RSLs de eficácia, de significado, de prevalência e incidência, de associação (etiologia), de precisão de testes de diagnóstico, de métodos mistos, de revisões de revisões e de *scoping* e de termos considerado algumas especificidades sobre a síntese dos dados iremos, de seguida: i) apresentar dados e uma discussão sobre a qualidade da evidência e risco de viés relativamente a uma RSL que integra estudos quantitativos e qualitativos e a uma revisão de revisões que temos vindo a desenvolver, ii) tecer algumas considerações sobre os níveis de evidência e graus de recomendação para a prática, muitas vezes confundidos com a avaliação crítica e qualidade dos estudos.

Nas RSLs em apreço, a pesquisa bibliográfica foi realizada nas bases de dados relevantes procurando informação publicada num intervalo de tempo específico, devidamente justificado em cada protocolo e relatório de revisão, seguindo a metodologia preconizada pelo JBI. Os artigos selecionados pela pesquisa bibliográfica nas bases de dados foram avaliados por dois revisores independentes relativamente aos títulos, aos resumos e aos textos completos para verificar a sua elegibilidade em relação aos critérios de inclusão definidos no protocolo de revisão. Os artigos que cumpriam os referidos critérios foram, de seguida, submetidos à avaliação crítica igualmente realizada por dois revisores. Em todo o processo as divergências que surgiram entre os revisores foram resolvidas através de discussão e com a participação de um terceiro revisor. Os dados foram extraídos e sintetizados tendo em conta a especificidade de cada revisão, como referido anteriormente, também por dois revisores.

Numa revisão em desenvolvimento (Apóstolo et al., 2013) com o objetivo de identificar e sintetizar as melhores evidências disponíveis sobre a eficácia das intervenções não farmacológicas de enfermagem no conforto de utentes com cancro, e sobre as suas

experiências de conforto relativamente àquelas intervenções, um total de 1413 estudos potencialmente relevantes foram identificados na pesquisa bibliográfica. Destes 274 eram duplicados. Dos restantes 1139 artigos, 827 foram excluídos após a leitura do título e análise do resumo. Dos 312 artigos restantes, 270 foram excluídos após a análise de texto completo por não cumprirem aos critérios de inclusão. A qualidade metodológica dos restantes 41 estudos foi avaliada. Finalmente, 13 artigos originais, num total de 1011 doentes com cancro, foram incluídos nesta revisão, 11 quantitativos (sete ERCs e cinco estudos *antes-após* - com um só grupo) e dois qualitativos. Os estudos quantitativos foram submetidos à avaliação da qualidade metodológica pela aplicação da Instrumento de Avaliação Crítica para Ensaios Clínicos Randomizados ou Quase-randomizados (JBI, 2014a) e os estudos qualitativos foram submetidos à avaliação crítica utilizando o Instrumento de Avaliação Crítica para Estudos Qualitativos (JBI, 2014a).

Qualidade dos estudos quantitativos

Como se pode observar na tabela 1, os sete ERCs incluídos apresentam 100% de respostas *sim* aos itens 6, 7, 8, 9 e 10. Nos restantes itens a percentagem de respostas, *sim* no conjunto dos artigos incluídos, oscila entre 0% (item 1) e 86% (item 3). Nenhum artigo apresenta resposta *sim* a todos os itens. O número de respostas *sim* oscila entre três (dois artigos) e nove (dois artigos). Entre as fragilidades metodológicas identificadas, com maior frequência, elencamos as seguintes:

Verificado um potencial viés de seleção em todos os artigos incluídos relacionado com a alocação dos participantes pelos grupos de tratamento, quer no que diz respeito ao processo de randomização, quer na ocultação do alocador ao processo. O procedimento de randomização dos participantes pelos grupos foi pseudoaleatória em três artigos e em quatro artigos a informação foi pouco clara. Apesar de todos eles relatarem ter utilizado algoritmos de computador para alocar aleatoriamente os participantes aos grupos, não está claro se esses algoritmos geram números verdadeiramente aleatórios. Como referimos anteriormente, os verdadeiros números aleatórios são gerados a partir de uma fonte de entropia, sendo, assim, totalmente imprevisíveis, enquanto os números pseudoaleatórios não dependem de uma fonte de entropia e são previsíveis.

Os estudos incluídos na revisão também foram alvos de viés de desempenho e de deteção. Devido ao desenho do estudo, era frequentemente impossível cegar os participantes ao tratamento a que estavam alocados. Apenas três estudos indicaram claramente o procedimento seguido. Num, o tratamento com placebo foi usado para controlo do viés de desempenho; noutra, há informação clara de que os participantes não foram informados se estavam incluídos no grupo experimental ou no grupo de controlo; no terceiro estudo há falta de informação detalhada sobre este procedimento.

Verificou-se também que raramente era ocultada aos avaliadores a informação sobre a que grupo pertenciam os participantes avaliados tendo esta sido apresentada somente em dois estudos.

A falta de descrição e inclusão na análise dos resultados de pessoas que não completaram o tratamento foi outra possível fonte de viés em três estudos. Somente três estudos apresentaram informação clara sobre este procedimento referindo ter usado a análise estatística ITT. No entanto, a descrição das variáveis de resultado incluíram apenas os utentes que haviam concluído o estudo.

Foram detetadas também situações em que a utilização dos testes para as análises estatísticas não era suficientemente clara. Tal aconteceu, por exemplo, nas situações em que foram utilizados testes paramétricos em amostras pequenas, sem ser feita referência ao cumprimento dos pressupostos relacionados com a utilização destes testes, como, por exemplo, a normalidade da distribuição.

A descrição da intervenção de controlo carece de descrição pormenorizada, quer em relação ao conteúdo do tratamento providenciado, quer em relação à frequência ou duração das sessões por ele abrangidas. Para descrever a intervenção de controlo, os autores destes ERCs recorrem, muitas vezes, a afirmações do tipo *no grupo de controlo foi usado tratamento usual ou tratamento padrão*. A falta destes pormenores impossibilita a comparação de estudos efetuados em diferentes contextos socioculturais uma vez que o tratamento usual é diferente nos diferentes contextos.

Tabela 1

Avaliação da qualidade metodológica dos ERCs incluídos

Citação	1	2	3	4	5	6	7	8	9	10
Adamsen et al., 2009	U	N	Y	Y	N	Y	Y	Y	Y	Y
Andersen et al., 2013	U	N	Y	N	Y	Y	Y	Y	Y	Y
Billhult, et al., 2007	N	N	Y	Y	N	Y	Y	Y	Y	Y
Catlin & Taylor-Ford, 2011	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Jane et al., 2011	U	Y	Y	Y	U	Y	Y	Y	Y	Y
Liu et al., 2008	N	Y	N	N	N	Y	Y	Y	Y	Y
Smith et al., 2001	U	U	Y	N	U	Y	Y	Y	Y	Y
%	0%	43%	86%	57%	29%	100%	100%	100%	100%	100%

Nota: Y = sim; N = não; U = não claro. Adaptado de “Non-pharmacological nursing interventions for the comfort of patients with cancer: A comprehensive systematic review” por João Apóstolo, Daniela Cardoso, Adriana Coelho, Elzbieta Bobrowicz-Campos. Manuscrito em preparação.

Os 16 artigos excluídos foram publicados em revistas de reconhecido mérito internacional com fatores de impacto no *Journal of Citation Reports* referentes ao ano 2015 entre 0,659 e 20,982.

Qualidade dos estudos qualitativos

Como se pode observar na tabela 2, os dois estudos incluídos apresentam 100% de respostas *sim* aos itens 9 e 10. Foi considerado, pelos revisores, que apresentavam processos adequados de recolha de dados e análise de dados, e as conclusões tiradas pelos investigadores eram baseadas nos dados recolhidos. Contudo, relativamente aos aspetos éticos, um dos estudos mencionou apenas a existência de consentimento informado dos participantes.

Nos restantes itens a percentagem de respostas *sim* nos dois artigos incluídos é de 0 (zero). As respostas a estes itens são maioritariamente consideradas como *não claras*. A qualidade metodológica dos estudos qualitativos é baixa sendo difícil aferir da confiabilidade e credibilidade, transferibilidade e confirmabilidade. Não é apresentada informação que globalmente permita ajuizar sobre a congruência onto-epistemológica, metodológica e posicionamento do investigador face à investigação. As fragilidades metodológicas identificadas são as seguintes:

Verificou-se estar em falta uma declaração clara sobre as premissas filosóficas subjacentes ao estudo, bem como, informação detalhada sobre a abordagem metodológica, o que reduziu substancialmente a confiabilidade dos resultados por eles reportados. Em ambos os estudos a informação sobre o passado cultural e teórico dos autores não foi fornecida. Além disso, nenhum dos dois estudos abordou a relação entre os investigadores e os participantes no momento da colheita de dados, e, em ambos, as vozes dos participantes foram relatadas de uma forma muito limitada, tornando as conclusões menos credíveis. A transferibilidade destes achados também é fraca, uma vez que os pontos de vista dos participantes e as experiências sobre as intervenções não farmacológicas de enfermagem foram extraídos, apenas, de dois estudos incluídos.

Apesar das fragilidades destes dois estudos, foram relatadas as suas conclusões e discutidas estas limitações, uma vez que esta é a melhor evidência disponível para compreender as experiências de conforto de utentes submetidos a intervenções não farmacológicas de enfermagem em contexto ambulatório e hospitalar.

Tabela 2

Avaliação da qualidade metodológica dos estudos qualitativos incluídos

Citação	1	2	3	4	5	6	7	8	9	10
Dunwoody et al., 2002	U	U	U	U	U	N	U	U	Y	Y
Liu et al., 2008	U	U	U	U	U	N	U	Y	Y	Y
%	0%	0%	0%	0%	0%	0%	0%	50%	100%	100%

Nota: Y = sim; N = não; U = não claro. Adaptado de "Non-pharmacological nursing interventions for the comfort of patients with cancer: A comprehensive systematic review" por João Apóstolo, Daniela Cardoso, Adriana Coelho, Elzbieta Bobrowicz-Campos. Manuscrito em preparação.

Numa revisão de revisões que desenvolvemos recentemente, sobre precisão de testes de diagnóstico relativamente à fragilidade na pessoa idosa, analisou-se a potencial inclusão de RSLs quantitativas, análises combinadas e meta-análises (Apóstolo et al., 2017). Os textos selecionados neste processo foram submetidos à avaliação crítica por dois revisores, utilizando o Instrumento de Avaliação Crítica para Revisões Sistemáticas e Síntese de Investigação (JBI, 2014d).

Neste processo, um total de 420 estudos potencialmente relevantes foram identificados na pesquisa bibliográfica. Destes, 75 eram duplicados. Dos restantes 345 textos, 325 foram excluídos após a avaliação do título e do resumo, e, em seguida, 10 foram excluídos após a análise de texto completo. Os restantes 10 estudos foram considerados como tendo critérios de inclusão e submetidos à avaliação da qualidade metodológica.

A fim de garantir a qualidade da evidência analisada, foi aplicado um ponto de corte para a inclusão de RSLs. Decidiu considerar-se como obrigatórios três itens do referido Instrumento de Avaliação Crítica para Revisões Sistemáticas e Síntese de Investigação (JBI, 2014d): item 2 (relacionado com a adequação dos critérios de inclusão para a pergunta de revisão), item 5 (relacionado com a adequação dos critérios utilizados para a avaliação crítica dos estudos incluídos) e item 6 (que verifica se a avaliação crítica foi realizada por dois ou mais revisores de forma independente). A escolha destes três itens está relacionada com a necessidade de evitar a inclusão das RSLs que não considerassem o risco de viés nos estudos primários ou que eram propensas ao viés de seleção devido à falta de controlo da subjetividade no processo de avaliação crítica dos estudos primários e falta de critérios de inclusão apropriados. Assim, as RSLs com resposta negativa a qualquer um dos três critérios de qualidade referidos foram excluídas. Em caso de informação pouco clara foram contactados os autores da revisão, a fim de obter esclarecimento. Após esta análise, cinco RSLs foram excluídas e cinco RSLs foram incluídas na revisão das revisões.

Qualidade das revisões avaliadas

Nas cinco revisões incluídas era apresentada uma clara e explícita questão de revisão (item 1), o processo de pesquisa nas fontes era adequado (item 4), foram usados critérios adequados para a avaliação crítica dos estudos primários (item 5), as recomendações, para a política e/ou prática, eram suportadas pelos dados relatados (item 10), e eram indicadas orientações específicas apropriadas para nova investigação (item 11). Numa revisão os critérios de inclusão não eram suficientemente detalhados para decidir se eram adequados ou não para a pergunta de revisão, sendo essa questão avaliada como *não clara* (item 2). Uma resposta *não clara* foi também considerada em relação à adequação da estratégia de pesquisa (item 3). Outra revisão forneceu informações insuficientes em relação ao processo de avaliação crítica, não sendo claro se a sua realização foi feita por dois ou mais revisores independentes (item 6). Falta de informação suficiente foi também observada em relação ao processo de extração de dados em três revisões que não especificaram o método usado para minimizar os erros no referido processo (item 7). Uma revisão não forneceu informações precisas sobre as razões da escolha do método de combinar os estudos (item 8). Nenhuma das RSLs incluídas avaliou a probabilidade de viés de publicação (item 9). Os instrumentos de avaliação crítica de estudos primários incluídos nas RSLs que foram incluídas nesta revisão de revisões (Apóstolo et al., 2017) foram os seguintes: QUADAS-2, *Quality in Prognosis Studies* (QUIPS), *Assessment Scale for the Measurement Properties of Health Status Questionnaires* e uma lista de 27 critérios construída especificamente para o efeito.

A tabela 3 ilustra os resultados da avaliação da qualidade metodológica das revisões incluídas.

Tabela 3
Avaliação da qualidade metodológica das revisões incluídas

Review	1	2	3	4	5	6	7	8	9	10	11
[34] Clegg et al., 2014	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
[35] Drubbel et al., 2014	Y	Y	U	Y	Y	U	U	U	N	Y	Y
[36] Pialoux et al., 2012	Y	U	Y	Y	Y	Y	U	Y	N	Y	Y
[37] Carpenter et al., 2014	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y
[38] Vermeulen et al., 2011	Y	Y	Y	Y	Y	Y	U	Y	N	Y	Y
%	100	80	80	100	100	80	40	80	0	100	100

Nota: Y = sim; N = não; U = não claro. Adaptado de "Predicting risk and outcomes for frail older adults: an umbrella review of available frailty screening tools," por J. Apóstolo, R. Cooke, E. Bobrowicz-Campos, S. Santana, M. Marcucci, A. Cano, ... C. Holland, 2017, *JBI Database of Systematic Reviews and Implementation Reports*, 15(4), pp. 1154-1208.

É importante notar, ainda, que nenhuma RSL incluída pesquisou literatura cinzenta. A inclusão de literatura cinzenta pode ser, para além da realização dos testes de funil, uma das formas de limitar o viés de publicação. Também é importante destacar que os resultados dos estudos primários fornecidos pelas revisões incluídas nesta revisão de revisões eram, frequentemente, insuficientemente detalhados. Por exemplo, em duas das RSLs incluídas os autores apresentaram coeficientes de correlação ou valores de sensibilidade e especificidade, mas sem esclarecer a base estatística utilizada para este fim, o que levantou o problema da interpretação dos dados comunicados. Noutra RSL incluída foram fornecidos diferentes tamanhos de efeito para os resultados adversos em saúde, sem ser feita referência à magnitude da exposição a estes resultados, o que fez a conversão de dados para uma estatística uniforme e a sua posterior comparação impossível. É possível que esses detalhes também não tenham sido apresentados nos estudos primários, contudo, como no caso da revisão de revisões não se vai verificar os estudos primários, é impossível confirmar a origem desta limitação. A falta de informações detalhadas limitou as análises que poderiam ser realizadas, constituindo um outro ponto fraco desta revisão de revisões.

As cinco RSLs foram excluídas porque: numa RSL os critérios para a avaliação crítica eram inapropriados, tendo sido utilizada a *STrengthening the Reporting of OBservational studies in Epidemiology* (STROBE) somente para avaliar os resumos e não foram avaliados os textos completos; em duas RSLs não foi feita qualquer avaliação crítica; e em duas RSLs os critérios de inclusão não eram apropriados para a questão de revisão. Estas cinco RSLs excluídas foram publicadas em revistas de reconhecido mérito internacional com fatores de impacto no *Journal of Citation Reports* referentes ao ano 2015 entre 2,198 e 26,509.

Como descrito nos anteriores exemplos foram excluídos ERCs e RSLs publicados em revistas de elevado impacto e reconhecimento internacional, significando que este pressuposto não é ainda condição necessária para a qualidade desejada das publicações. Há, assim, necessidade de desenvolver formação, quer aos investigadores primários, quer aos revisores para que desenvolvam revisões de elevada qualidade. Mesmo nas revisões incluídas há, como referimos acima, algumas fragilidades que podem ser minimizadas durante o seu desenvolvimento a fim de melhorar a qualidade da síntese produzida.

Adicionalmente, às limitações anteriormente discutidas, há diversas RSLs publicadas em que os níveis de evidência são utilizados para classificar as revisões sem que seja feita avaliação crítica. Há autores que consideram, ainda, os estudos quantitativos com níveis de evidência elevados e os qualitativos com níveis de evidência mais baixos.

A JBI reconhece diferentes níveis de evidência para cada tipologia de revisões (JBI, 2013b). Na verdade, quando um nível de evidência é atribuído à informação sintetizada com base no seu desenho do estudo, o clínico que a consome deverá ser capaz de tecer

um juízo sobre a qualidade metodológica e rigor da evidência produzida. Apresentamos, a título de exemplo, os considerados níveis de evidência de eficácia e de significado.

Os níveis de evidência para estudos de eficácia têm os seguintes níveis: 1.a - RSL de ensaios clínicos randomizados; 1.b - RSL de ensaios clínicos randomizados e outros desenhos de estudo; 1.c - ensaios clínicos randomizados; 1.d - pseudo-ensaios clínicos randomizados; 2 - desenhos quasi-experimentais; 2.a - RSL de estudos quasi-experimentais; 2.b - RSL de estudos quasi-experimentais e outros desenhos de estudos de nível inferior; 2.c - estudo quasi-experimental controlado prospectivamente; 2.d - pré-teste-pós-teste ou estudo histórico/retrospectivo com grupo de controle.

Os níveis de evidência para estudos de significado, que incluem os estudos qualitativos, têm os seguintes cinco níveis: 1 - RSLs de estudos qualitativos ou métodos mistos; 2 - sínteses qualitativas ou com métodos mistos; 3 - estudo individual qualitativo; 4 - RSL sobre opinião/parecer de peritos; 5 - parecer do perito.

É importante notar que esses níveis de evidência fornecem um ranking baseado na provável melhor evidência disponível, e que não devem ser tomados como uma medida definitiva da melhor evidência disponível. Pode acontecer que a evidência sintetizada a partir de estudos observacionais, no final da avaliação, é preferível à que provém de ERCs (JBI, 2014f).

Os níveis de evidência JBI para as intervenções, diagnóstico e prognóstico estão alinhados com a abordagem *Grading of Recommendations Assessment, Development and Evaluation* (GRADE) pré-classificando os resultados com base no desenho do estudo, que são, em seguida, atualizados (sobem ou descem), dependendo de um número de fatores como a avaliação crítica e o risco de viés, o tamanho do efeito, relações dose-resposta, e outros fatores confundentes (Andrews, Guyatt et al. 2013; Andrews, Schünemann et al., 2013; JBI, 2014e; Apóstolo & Cardoso, 2014).

Em relação aos estudos qualitativos e opinião de peritos não é apropriado a distinção dos diferentes desenhos de estudos através de uma hierarquia. Assim, numa escala de classificação, que vai de alta a muito baixa, todos os estudos qualitativos são pré-classificados com um nível de evidência *alta* enquanto os textos de opinião de peritos são pré-classificados com um nível de evidência *baixa*. Esta pré-classificação pode ser reduzida tendo em conta o grau de credibilidade e de fiabilidade dos estudos (JBI, 2014g).

Simultaneamente, a JBI atribui um grau de recomendação para a prática para apoiar os profissionais de saúde na implementação da evidência na prática. Os novos graus de recomendação JBI (informadas pela GRADE) são um sistema binário, com duas opções: *forte* ou *fraco*. A recomendação *forte* é representada pela letra A e a recomendação *fraco* é representado pela letra B. Este sistema pode ser facilmente interpretado por clínicos e utentes (JBI, 2013b; 2014b).

CONSIDERAÇÕES FINAIS

Os diferentes modelos descritos na literatura referem diferentes passos para implementação da ciência, considerando a produção e síntese crítica do conhecimento, a sua disseminação e a sua implementação no contexto das organizações, e, finalmente, a avaliação do impacto nos utentes e organizações. A geração de evidências inicia-se com a produção de investigação primária - empírica - rigorosa, utilizando os mais diversos desenhos de investigação que lhe conferem credibilidade que são depois sujeitos a um processo de síntese.

A RSL é uma investigação secundária complexa, detalhada e reprodutível, que envolve um compromisso significativo de tempo e outros recursos. Nesta metodologia é utilizado um processo para localizar e agregar a partir da literatura primária todas as evidências existentes num determinado tópico, usando uma fonte secundária de dados. A transferência das evidências é a disponibilização da evidência sintetizada aos profissionais de saúde, às unidades e aos sistemas de saúde a nível mundial, através de revistas científicas, outras publicações, educação e treino, e também através do suporte à decisão. Inclui o desenho e implementação de métodos para organizar e transferir a informação de forma compreensível e utilizável, podendo esta ser integrada em sistemas de informação para apoio à decisão clínica.

Para que a evidência se torne facilmente acessível aos seus utilizadores e consumida, a informação resultante das revisões sistemáticas pode ainda ser resumida e condensada numa variedade de formatos, eletrónicos ou impressos, como sumários de evidência, recomendações baseadas em evidência, folhas de informação de melhores práticas, folhas de informação ao consumidor e infográficos.

A implementação da evidência é um processo complexo, definido como um conjunto de atividades de capacitação destinadas a envolver os principais interessados para que a tomada de decisão possa ser informada pela evidência e gerar melhoria sustentada na qualidade da prestação de cuidados de saúde.

Um utilizador esclarecido tem necessidade de aperfeiçoar as suas capacidades de análise, síntese e utilização de evidência científica. Aperfeiçoar estas capacidades pressupõe treino para questionar as práticas clínicas em uso e pesquisar a melhor informação científica que está disponível em plataformas e bases de dados científicas. Pressupõe, ainda, o desenvolvimento de capacidades que permitam ajuizar criticamente sobre a qualidade da informação pesquisada e o conhecimento de metodologias para a utilização dessa evidência na prática.

AGRADECIMENTOS

O autor agradece o apoio na revisão do texto e tradução dos instrumentos de avaliação crítica de Daniela Cardoso (RN), Elzbieta Bobrowicz-Campos (PhD), Paulo Costa (RN) e Rogério Rodrigues (PhD).

Agradece à Mónica Jardim (PhD) pelo apoio na edição e revisão de texto.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aarons, G. (2006). Transformational and transactional leadership: Association with attitudes toward evidence-based practice. *Psychiatric Services*, 57(8), 1162-1169. doi: 10.1176/appi.ps.57.8.1162
- Aarons, G. A., Ehrhart, M. G., & Farahnak, L. R. (2014). The implementation leadership scale (ILS): development of a brief measure of unit level implementation leadership. *Implementation Science*, 9(1), 45. doi: 10.1186/1748-5908-9-45
- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health*, 38(1), 4-23. doi: 10.1007/s10488-010-0327-7
- Agency for Healthcare Research and Quality. (2001). *Accelerating change and translating research into practice (TRIP)-II: Fact Sheet*. Recuperado de: <http://archive.ahrq.gov/research/findings/factsheets/translating/tripfac/trip2fac.html>
- Andrews, J., Guyatt, G., Oxman, A. D., Alderson, P., Dahm, P., Falck-Ytter, Y., ... & Brozek, J. (2013). GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *Journal of clinical epidemiology*, 66(7), 719-725.
- Andrews, J. C., Schünemann, H. J., Oxman, A. D., Pottie, K., Meerpohl, J. J., Coello, P. A., ... & Elbarbary, M. (2013). GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. *Journal of clinical epidemiology*, 66 (7), 726-735.
- Anney, V. N. (2014). Ensuring the quality of the findings of qualitative research: looking at trustworthiness criteria. *Journal of Emerging Trends in Educational Research and Policy Studies*, 5(2), 272-281.

- Apóstolo, J., & Cardoso, D. (2014). Systematic literature review: From the levels of evidence to the grades of recommendation. *Aten Primaria*, 46(Espec Cong 1), 3.
- Apóstolo, J., & Gameiro, M. (2005). Referências onto-epistemológicas e metodológicas da investigação em enfermagem: uma análise crítica. *Revista Referência*, 2(1), 29-38.
- Apóstolo, J., Bobrowicz-Campos, E., Rodrigues, M., Castro, I., & Cardoso, D. (2016). The effectiveness of non-pharmacological interventions in older adults with depressive disorders: A systematic review. *International Journal of Nursing Studies*, 58(6), 59-70. doi: 10.1016/j.ijnurstu.2016.02.006
- Apóstolo, J., Cardoso, D., & Rodrigues, M. A. (2016). It takes three to tango: embracing EBP. *JBI Database of Systematic Reviews and Implementation Reports*, 14(4), 1-2. doi: 10.11124/JBISIRIR-2016-288
- Apóstolo, J., Cooke, R., Bobrowicz-Campos, E., Santana, S., Marcucci, M., Cano, A., ... Holland, C. (2015). Predicting risk and outcomes for frail older adults: a protocol for an umbrella review of available frailty screening tools. *JBI Database of Systematic Reviews and Implementation Reports*, 13(12), 14-24. doi: 10.11124/jbisirir-2015-2468
- Apóstolo, J., Cooke, R., Bobrowicz-Campos, E., Santana, S., Marcucci, M., Cano, A., ... Holland, C. (2016). Effectiveness of the interventions in preventing the progression of pre-frailty and frailty in older adults: a systematic review protocol. *JBI Database of Systematic Reviews and Implementation Reports*, 14(1), 4-19. doi: 10.11124/jbisirir-2016-2467.
- Apóstolo, J., Cooke, R., Bobrowicz-Campos, E., Santana, S., Marcucci, M., D'Avanzo, B; Cano, A., ... Holland, C. (In press). Effectiveness of the interventions in preventing the progression of pre-frailty and frailty in older adults: a systematic review. *JBI Database of Systematic Reviews and Implementation Reports*.
- Apóstolo, J., Cooke, R., Bobrowicz-Campos, E., Santana, S., Marcucci, M., Cano, A., ... Holland, C. (2017). Predicting risk and outcomes for frail older adults: an umbrella review of available frailty screening tools. *JBI Database of Systematic Reviews and Implementation Reports*, 15(4), 1154-1208. doi: 10.11124/JBISIRIR-2016-003018
- Apóstolo, J., Queirós, P., Rodrigues, M., Castro, I., & Cardoso, D. (2015). The effectiveness of nonpharmacological interventions in older adults with depressive disorders: a systematic review. *JBI Database of Systematic Reviews and Implementation Reports* 13(6), 220-278. doi: 10.11124/jbisirir-2015-1718
- Apóstolo, J., Mendes, A., Bath-Hextall, F., Rodrigues, R., Santos, J., & Cardoso, D. (2013). The use of non-pharmacological nursing interventions on the comfort of cancer patients: A comprehensive systematic review protocol. *JBI Database of Systematic Reviews and Implementation Reports*, 11(2), 372-388.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. doi: 10.1080/1364557032000119616
- Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P. (2015). Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *International journal of evidence-based healthcare*, 13(3), 132-140.

- Campbell, J. M., Klugar, M., Ding, S., Carmody, D. P., Hakonsen, S. J., Jadotte, Y. T., ... Munn, Z. (2015). Diagnostic test accuracy: methods for systematic review and meta-analysis. *International Journal of Evidence-Based Healthcare*, 13(3), 154-162.
- Chan, A. W. (2008). Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS Med*, 5(11), e230. doi:10.1371/journal.pmed.0050230
- Coelho, A., Parola, V., Cardoso, D., Escobar, M., & Apóstolo, J. (2016). The use of non-pharmacological interventions for the comfort of patients in palliative care: a scoping review protocol. *JBI Database of Systematic Reviews and Implementation Reports*, 14(2), 64-77. doi: 10.11124/jbisrir-2016-2440
- Cooksey, D. (2006). *A review of UK health research funding*. Recuperado de https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/228984/0118404881.pdf
- Crandell, J. L., Voils, C. I., Chang, Y-K., & Sandelowski, M. (2011). Bayesian data augmentation methods for the synthesis of qualitative and quantitative research findings. *Quality and Quantity*, 45(3), 653-669.
- Daniel, W. W. (1999) *Biostatistics: A Foundation for Analysis in the Health Sciences* (7th ed.). New York: John Wiley & Sons.
- Dawes, M., Summerskill, W., Glasziou, P., Cartabellotta, A., Martin, J., Hopayian, K., ... Osborne, J. (2005). Sicily statement on evidence-based practice. *BMC medical education*, 5(1), 1.
- de Smith, M. J. (2015). *STATSREF: Statistical Analysis Handbook - a web-based statistics*. Recuperado de: <http://www.statsref.com/StatsRefSample.pdf>
- Direção Geral da Saúde. (2012). *Plano Nacional de saúde 2012-2016*. Recuperado de: <http://pns.dgs.pt/pns-versao-completa/>
- Dubben, H. H., & Beck-Bornholdt, H. P. (2005). Systematic review of publication bias in studies on publication bias. *The BMJ*, 331(7514), 433-434. doi: 10.1136/bmj.38478.497164.F7
- Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS ONE*, 8(7), e66844. doi: 10.1371/journal.pone.0066844
- Ehrhart, M. G., Aarons, G. A., & Farahnak, L. R. (2014). Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). *Implementation Science*, 9, 157. doi: 10.1186/s13012-014-0157-1
- Ehrhart, M., Schneider, B., & Macey, W. (2014). *Organizational climate and culture: An introduction to theory, research, and practice*. New York: Routledge.
- Escopo. (n.d.). In *Dicionário Priberam da Língua Portuguesa*. Recuperado de <http://www.priberam.pt/dlpo/escopo>
- Foley, L. (2001). *Analysis of an on-line random number generator*. Recuperado de <https://www.random.org/analysis/Analysis2001.pdf>
- Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., ... McBurnie, M. A. (2001). Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences*, 56(3), M146-M156.

- Gagnier, J. J., Kienle, G., Altman, D. G., Moher, D., Sox, H., & Riley, D. (2013). The CARE Guidelines: Consensus-Based Clinical Case Reporting Guideline Development. Headache. *The Journal of Head and Face Pain*, 53(10), 1541-1547
- Gifford, W., Davies, B., Edwards, N., Griffin, P., & Lybanon, V. (2007). Managerial leadership for nurses' use of research evidence: An integrative review of the literature. *Worldviews on Evidence-Based Nursing*, 4(3), 126-145.
- Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, 2(3), 109-112. doi: 10.4103/2229-3485.83221
- Hannes, K., & Lockwood, C. (2011). Pragmatism as the philosophical foundation for the Joanna Briggs meta-aggregative approach to qualitative evidence synthesis. *Journal of Advanced Nursing*, 67(7), 1632-1642. doi: 10.1111/j.1365-2648.2011.05636.x
- Held, D. (1980). *Introduction to critical theory: Horkheimer to Habermas*. Berkeley: University of California Press.
- Higgins, J. P., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions, version 5.1.0* [updated March 2011]. Recuperado de: www.cochrane-handbook.org.
- Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ*, 327, 557-560.
- Hoffman, M. (1989). Critical Theory and the Inter-paradigm Debate. In H. C. Dyer & L. Mangasarian (Eds.), *The Study of International Relations: The State of the Art* (pp. 60-86). London: Palgrave Macmillan UK.
- International Council of Nurses. (2012). *Closing the gap: From evidence to action*. Geneva: Author.
- Joanna Briggs Institute. (2013a). *Handbook for evidence transfer centres*. Recuperado de: http://joannabriggs.org/assets/docs/jbc/operations/connectNodeHankbook/NodeCentresDraft_MW2.pdf
- Joanna Briggs Institute. (2013b). *New JBI Levels of Evidence: Developed by the Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party October 2013*. Recuperado de: http://joannabriggs.org/assets/docs/approach/JBI-Levels-of-evidence_2014.pdf
- Joanna Briggs Institute. (2014a). *Joanna Briggs Institute reviewers' manual: 2014 edition*. Recuperado de: <http://joannabriggs.org/assets/docs/sumari/reviewersmanual-2014.pdf>
- Joanna Briggs Institute. (2014b). *Joanna Briggs Institute reviewers' manual: 2014 edition/ supplement. The systematic review of prevalence and incidence data*. Recuperado de: http://joannabriggs.org/assets/docs/sumari/ReviewersManual_2014-The-Systematic-Review-of-Prevalence-and-Incidence-Data_v2.pdf
- Joanna Briggs Institute. (2014c). *Joanna Briggs Institute reviewers' manual: 2014 edition/supplement. Methodology for JBI mixed methods systematic reviews*. Recuperado de: http://joannabriggs.org/assets/docs/sumari/ReviewersManual_Mixed-Methods-Review-Methods-2014-ch1.pdf
- Joanna Briggs Institute. (2014d). *Joanna Briggs Institute reviewers' manual: 2014 edition/ supplement. Methodology for JBI umbrella reviews*. Recuperado de: http://joannabriggs.org/assets/docs/sumari/ReviewersManual-Methodology-JBI_Umbrella%20Reviews-2014.pdf

- Joanna Briggs Institute. (2014e). *The Joanna Briggs Institute levels of evidence and grades of recommendation working party*. Supporting document for the Joanna Briggs Institute levels of evidence and grades of recommendation*. Recuperado de: <http://joannabriggs.org/assets/docs/approach/Levels-of-Evidence-SupportingDocuments.pdf>
- Joanna Briggs Institute. (2014f). *The Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party. Supporting Document for the Joanna Briggs Institute Levels of Evidence and Grades of Recommendation*. <http://joannabriggs.org/assets/docs/approach/Levels-of-Evidence-SupportingDocuments-v2.pdf>
- Joanna Briggs Institute. (2014g). *The Joanna Briggs Institute Levels of Evidence and Grades of Recommendation Working Party. Summary of Findings Tables for Joanna Briggs Institute Systematic Reviews*. Recuperado de http://joannabriggs.org/assets/docs/sumari/ReviewersManual-2014-Summary-of-Findings-Tables_v2.pdf
- Joanna Briggs Institute. (2015a). *Discussion paper: The JBI model of evidence-informed healthcare: A model reconsidered*. Recuperado de: http://joannabriggs.org/assets/docs/approach/The_JBI_Model_of_Evidence_-_Healthcare-A_Model_Reconsidered.pdf
- Joanna Briggs Institute. (2015b). *Joanna Briggs Institute reviewers' manual: 2015 edition/supplement. The systematic review of studies of diagnostic test accuracy*. Recuperado de: http://www.joannabriggs.org/assets/docs/sumari/Reviewers-Manual_The-systematic-review-of-studies-of-diagnostic-test-accuracy.pdf
- Joanna Briggs Institute. (2015c). *Joanna Briggs Institute reviewers' manual: 2015 edition/supplement. Methodology for JBI scoping reviews*. Recuperado de: http://joannabriggs.org/assets/docs/sumari/Reviewers-Manual_Methodology-for-JBI-Scoping-Reviews_2015_v2.pdf
- Joanna Briggs Institute. (2016). *Joanna Briggs Institute Reviewers' Manual: 2016 edition*. Australia: The Joanna Briggs Institute.
- Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. *Systematic Reviews*, 1(1), 10. doi: 10.1186/2046-4053-1-10
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.
- Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *International journal of evidence-based healthcare*, 13(3), 179-187.
- Lomba, L., Kroll, T., Apóstolo, J. M., Gameiro, M., & Apóstolo, J. (2016). The use of motivational interviews by nurses to promote health behaviors in adolescents: A scoping review protocol. *JBI Database of Systematic Reviews and Implementation Reports*, 14(5), 27-37. doi: 10.11124/jbisrir-2016-002564
- Melnik, B. M., & Newhouse, R. (2014). Evidence-based practice versus evidence informed practice: A debate that could stall forward momentum in improving healthcare quality, safety, patient outcomes and costs. *Worldviews on Evidence-Based Nursing*, 11(6), 347-349. doi: 10.1111/wvn.12070
- McArthur, A., Klugarova, J., Yan, H., & Florescu, S. (2015) Innovations in the systematic review of text and opinion. *Int J Evid Based Healthc*, 13:188-195. doi:10.1097/XEB.0000000000000060.

- Melnik, B., Gallagher-Ford, L., Long, L., & Fineout-Overholt, E. (2014). The establishment of evidence-based practice competencies for practicing registered nurses and advanced practice nurses in real-world clinical settings: proficiencies to improve healthcare quality, reliability, patient outcomes, and costs. *Worldviews on Evidence-Based Nursing*, 11(1), 5-15. doi: 10.1111/wvn.12021
- Moola, S., Munn, Z., Sears, K., Sfetcu, R., Currie, M., Lisy, K., ... Mu, P. (2015). Conducting systematic reviews of association (etiology): The Joanna Briggs Institute's approach. *International Journal of Evidence-Based Healthcare*, 13(3), 163-169.
- Munn, Z., Moola, S., Lisy, K., Riitano, D., & Tufanaru, C. (2015). Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. *International Journal of Evidence-Based Healthcare*, 13(3), 147-153.
- Naing, L., Winn T., & Rusli, B. N. (2006). Practical issues in calculating the sample size for prevalence studies. *Archives of Orofacial Sciences*, 1, 9-14.
- Noblit, G., & Hare, R. (1988). *Meta-ethnography: synthesizing qualitative studies*. Newbury Park, CA: Sage.
- Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research*, 14, 2. doi: 10.1186/1472-6963-14-2
- Ordem dos Enfermeiros. (2006). *Investigação em enfermagem. Tomada de posição*. Recuperado de: http://www.ordemenfermeiros.pt/tomadasposicao/Documents/TomadaPosicao_26Abr2006.pdf
- Ordem dos Enfermeiros. (2012). *Padrões de qualidade dos cuidados de enfermagem – enquadramento conceptual. Enunciados descritivos*. Recuperado de: <http://www.ordemenfermeiros.pt/publicacoes/Documents/divulgar%20-%20padroes%20de%20qualidade%20dos%20cuidados.pdf>
- Parola, V., Coelho, A., Cardoso, D., Blanco-Blanco, J., Gea-Sanchez, M., & Apóstolo, J. (2016). The prevalence of burnout in health professionals working in palliative care: A systematic review protocol. *JBI Database of Systematic Reviews and Implementation Reports*, 14(3), 45-50.
- Pearson, A., Field, J., & Jordan, Z. (2007). *Evidence-based clinical practice in nursing and health care: Assimilating research, experience and expertise*. Oxford: Blackwell Publishing.
- Pearson, A., Jordan, Z., & Munn, Z. (2012). Translational science and evidence-based healthcare: a clarification and reconceptualization of how knowledge is generated and used in healthcare. *Nursing Research and Practice*, 2012. doi: 10.1155/2012/792519
- Pearson, A., Weeks, S. & Stern, C. (2011). *Translation Science and the JBI model of evidence-based healthcare*. Philadelphia: Lippincott, Williams & Wilkins.
- Pearson, A., White, H., Bath-Hextall, F., Salmond, S., Apóstolo, J., Kirkpatrick, P., & Lockwood, C. (2014). *A mixed methods approach to evidence synthesis*. Philadelphia: Lippincott, Williams & Wilkins.
- Pearson, A., White, H., Bath-Hextall, F., Salmond, S., Apóstolo, J., & Kirkpatrick, P. (2015). A mixed-methods approach to systematic reviews. *International Journal of Evidence-Based Healthcare*, 13(3), 121-131. doi: 10.1097/XEB.0000000000000052

- Pearson, A., Wiechula, R., Court, A., & Lockwood, C. (2005). The JBI model of evidence-based healthcare. *International Journal of Evidence-Based Healthcare*, 3, 207-215.
- Pedreiro, A. T., Loureiro, H., Ferreira, M., Cardoso, D., & Apóstolo, J. (2016). Measuring and evaluating adjustment to retirement: a scoping review protocol. *JBI Database of Systematic Reviews and Implementation Reports*, 14(9), 79-84. doi: 10.111124/JBISRIR-2016-003079
- Pijpers, E., Ferreira, I., Stehouwer, C., & Nieuwenhuijzen Kruseman, A. C. (2012). The frailty dilemma. Review of the predictive accuracy of major frailty scores. *European Journal of Internal Medicine*, 23(2), 118-123. doi: 10.1016/j.ejim.2011.09.003
- Popper, K. (1992). *Em busca de um mundo melhor*. Lisboa, Portugal: Fragmentos.
- Popper, K. (1993). *A lógica da pesquisa científica*. São Paulo, Brasil: Cultrix.
- Popper, K. (1997). *O conhecimento e o problema corpo-mente*. Lisboa, Portugal: Edições 70.
- Queirós, P., Santos, E., Apóstolo, J., Cardoso, D., Cunha, M., & Rodrigues, M. (2014). The effectiveness of cleansing solutions for wound treatment: A systematic review. *JBI Database of Systematic Reviews and Implementation Reports*, 12(10), 121-151.
- Rockwood, K., & Mitnitski, A. (2011). Frailty defined by deficit accumulation and geriatric medicine defined by frailty. *Clinics in Geriatric Medicine*, 27(1), 17-26.
- Rubenstein, L. Z., Stuck, A. E., Siu, A. L., & Wieland, D. (1991). Impacts of geriatric evaluation and management programs on defined outcomes: overview of the evidence. *Journal of the American Geriatrics Society*, 39(S1), 8S-16S. doi: 10.1111/j.1532-5415.1991.tb05927.x
- Sackett, D. L., Rosenberg, W., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *The BMJ*, 312(7023), 71-72.
- Sandelowski, M., Leeman, J., Knafl, K., & Crandell, J. L. (2013). Text-in-context: a method for extracting findings in mixed-methods mixed research synthesis studies. *Journal of Advanced Nursing*, 69(6), 1428-1437.
- Sandelowski, M., Voils, C. I., & Barroso, J. (2006). Defining and designing mixed research synthesis studies. *Research in the Schools*, 13(1), 29-40.
- Sandelowski, M., Voils, C. I., & Barroso, J. (2007). Comparability work and the management of difference in research synthesis studies. *Social Science & Medicine*, 64(1), 236-247.
- Santos, E., & Cunha, M. (2013). Interpretação crítica dos resultados estatísticos de uma meta-análise: Estratégias metodológicas. *Millenium*, 44 (janeiro/junho), 85-98.
- Schaffer, M. A., Sandau, K. E., & Diedrick, L. (2013). Evidence-based practice models for organizational change: overview and practical applications. *Journal of Advanced Nursing*, 69(5), 1197-1209.
- Schulz, K. F., & Grimes, D. A. (2002). Allocation concealment in randomised trials: defending against deciphering. *Lancet*, 359(9306), 614-618.
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75.

- Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559–1573. doi:10.1002/sim.1187
- Titler, M. G. (2008). The evidence for evidence-based practice implementation. In R. G. Hughes (ed.), *Patient safety and quality: An evidence-based handbook for nurses* (chapter 7). Recuperado de <http://www.ncbi.nlm.nih.gov/books/NBK2659/>
- Titler, M. G., & Everett, L. Q. (2001). Translating research into practice. Considerations for critical care investigators. *Critical Care Nursing Clinics of North America*, 13(4), 587-604.
- Tufanaru, C., Munn, Z., Stephenson, M., & Aromataris, E. (2015). Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. *International Journal of Evidence-Based Healthcare*, 13(3), 196-207.
- West, S. L., Gartlehner, G., Mansfield, A. J., Poole, Ch., Tant, E., Lenfestey, N., ... Lohr, K. (2010). *Comparative Effectiveness Review Methods: Clinical Heterogeneity*. Recuperado de: <http://www.ncbi.nlm.nih.gov/books/NBK53310/>
- Whiting, P., Rutjes, A. W., Reitsma, J. B., Bossuyt, P. M., & Kleijnen, J. (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology*, 3(1), 25.
- Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... & Bossuyt, P. M. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155(8), 529-536.
- Wise, N. J. (2009). Maintaining magnet status: establishing an evidence-based practice committee. *AORN Journal*, 90(2), 205-213.



IACS
UNIDADE DE INVESTIGAÇÃO
EM CIÊNCIAS DA SAÚDE
ENFERMAGEM



Escola Superior de
Enfermagem de Coimbra



UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR