



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

**ESCOLA DE MEDICINA E CIÊNCIAS DA VIDA
PROGRAMA DE PÓS-GRADUAÇÃO EM ODONTOLOGIA
ÁREA DE CONCENTRAÇÃO EM SAÚDE COLETIVA**

CAROLINE SOUZA DOS SANTOS

**PSYCHOMETRIC PROPERTIES OF FOOD FREQUENCY
QUESTIONNAIRE USED IN DENTISTRY: A SYSTEMATIC
REVIEW**

**PROPRIEDADES PSICOMÉTRICAS DE QUESTIONÁRIOS
DE FREQUÊNCIA ALIMENTAR USADOS NA
ODONTOLOGIA: UMA REVISÃO SISTEMÁTICA**

Curitiba

2022

CAROLINE SOUZA DOS SANTOS

**PSYCHOMETRIC PROPERTIES OF FOOD FREQUENCY
QUESTIONNAIRE USED IN DENTISTRY: A SYSTEMATIC
REVIEW**

**PROPRIEDADES PSICOMÉTRICAS DE QUESTIONÁRIOS
DE FREQUÊNCIA ALIMENTAR USADOS NA
ODONTOLOGIA: UMA REVISÃO SISTEMÁTICA**

**Dissertação apresentada ao Programa de
Pós-Graduação em Odontologia da
Pontifícia Universidade Católica do Paraná,
como parte dos requisitos para obtenção
do título de Mestre em Odontologia, Área
de Concentração em Multidisciplinaridades
em Saúde com Ênfase em Saúde Coletiva.**

Orientador: Prof. Dr. Juliana S. Rocha Orsi

Coorientador: Prof. Dr. Rodrigo N. Rached

Coorientador: Prof. Dr. Samuel J. Moysés

Curitiba

2022

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Pamela Travassos de Freitas – CRB 9/1960

Santos, Caroline Souza dos
S237p
2022 *Psychometric properties of food frequency questionnaire used in dentistry : a systematic review* = Propriedades psicométricas de questionários de frequência alimentar usados na odontologia : uma revisão sistemática / Caroline Souza dos Santos ; orientador: Juliana S. Rocha Orsi ; coorientadores: Rodrigo N. Rached, Samuel J. Moysés. – 2022.
111 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2022
Inclui bibliografias

1. Odontologia. 2. Ingestão de alimentos. 3. Comportamento alimentar. 4. Cáries dentárias. I. Rocha, Juliana Schaia. II. Rached, Rodrigo Nunes. III. Moyses, Samuel Jorge. IV. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Odontologia. V. Título.

CDD 21. ed. – 617.6



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

Escola de Ciências da Vida

Programa de Pós-Graduação em Odontologia

TERMO DE APROVAÇÃO

CAROLINE SOUZA DOS SANTOS

PROPRIEDADES PSICOMÉTRICAS DE QUESTIONÁRIOS DE FREQUÊNCIA ALIMENTAR USADOS NA ODONTOLOGIA: UMA REVISÃO SISTEMÁTICA

Dissertação apresentada ao Programa de Pós-Graduação em Odontologia da Pontifícia Universidade Católica do Paraná, como parte dos requisitos parciais para a obtenção do Título de **Mestre em Odontologia**, Área de Concentração em **Saúde Coletiva**.

Orientador(a):

Prof. Dr. Juliana Schaia Rocha Orsi
Programa de Pós-Graduação em Odontologia, PUCPR

Profª Dr. Renata Iani Werneck
Programa de Pós-Graduação em Odontologia, PUCPR

Profª Drª Márcia Helena Baldani Pinto
Programa de Pós-Graduação em Odontologia, UEPG

CURITIBA, 03 DE MARÇO DE 2022.

AGRADECIMENTOS

“Um sonho sonhado sozinho, é apenas um sonho. Um sonho sonhado juntos é o início de uma nova realidade”

(Yoko Ono)

Este é um momento muito especial, um sonho está se tornando realidade, agradeço a todos aqueles que compartilharam dessa etapa tão desejada e importante da minha vida pessoal e profissional.

Em primeiro lugar agradeço a **Deus**, por ter ouvido as minhas orações e que esteve comigo em todos os momentos me dando força, sabedoria e discernimento durante toda a minha caminhada.

Emocionada, não tenho palavras para expressar o meu profundo e sincero agradecimento a minha querida orientadora, **Juliana Schaia Rocha Orsi**, que com todo carinho, dedicação, ensinamentos e paciência me auxiliou a iniciar e trilhar esse caminho na pesquisa científica. Agradeço sua disponibilidade e atenção em me conduzir esta proposta de pesquisa. Me deu apoio em todos os momentos de dificuldade, fadiga, angústia, alegrias e conquistas. Agradeço a Deus por ter cruzado os nossos caminhos e que nos permita conviver e estar sempre juntas. Ouvir seus ensinamentos é sempre um grande aprendizado. À ela todo o meu carinho, amor, respeito e reconhecimento.

Um agradecimento especial ao meu co-orientador, **Rodrigo Nunes Rached**, que com sua disponibilidade e atenção me conduziu a bolsa de isenção do Centro Integrado de Pesquisa (CIP) da PUCPR, foi imprescindível para o desenvolvimento e finalização desta pesquisa. Ao professor e co-orientador, **Samuel Jorge Moysés** que com sua sabedoria, atenção, ideias e críticas construtivas, foram essenciais para o desenvolvimento dessa pesquisa. A estes agradeço que brilhantemente participou da minha qualificação da dissertação do mestrado.

Às professoras, **Marcia Baldani** e **Renata Iani Werneck**, que fazem parte desta conquista, gentilmente aceitaram fazer parte da minha banca, que mesmo de longe sempre tive grande afeto e admiração.

À **Coorte Materno Infantil de Curitiba** (COOSMIC), por todas as amizades construídas, conhecimento consolidados e experiências trocadas ficará marcado para sempre em meu coração.

Agradeço a **Pontifícia Universidade Católica do Paraná**, por me proporcionar uma formação completa e de qualidade.

A **coordenação, professores e colegas** do **PPGO** que me acolheram tão carinhosamente e oportunizaram aprender e transmitir meu conhecimento ao longo desses dois anos de estudo.

Á todos meus **amigos** que mesmo distantes participaram ativamente de todo o processo, compartilhando de todos os momentos alegres e angustiantes desta etapa da minha vida.

Á minha amada **FAMÍLIA**, meus pais **Isaura e Antônio**, ao meu esposo **Jonathan**, por todo amor, orações, paciência e motivação nesta trajetória da minha formação acadêmica. Ao meu irmão **Tiago** e minha cunhada **Glaucia**, por todo carinho, ensinamento e dedicação durante esse momento. Ao meu amado e desejado sobrinho **Davi Antônio** à espera de sua chegada tem sido a minha alegria, enchido o meu coração de amor e ternura. Vocês foram importantes para a realização deste trabalho, agradeço o exemplo e a motivação pela minha escolha profissional.

SUMÁRIO

ARTIGO - Versão em português	6
Resumo	7
Introdução	9
Métodos	11
Resultados	19
Discussão.....	26
Conclusão	32
Referências.....	33
MANUSCRIPT- English version	38
Abstract.....	39
Introduction.....	40
Methds	42
Results.....	50
Discussion.....	58
Conclusion	64
References	66
Figure Legends	70
APÊNDICES	71
Apêndice 1. Resultados do Teste de Wilcox para avaliação da existência de erro sistemático nas avaliações de qualidade	71
Apêndice 2. Conferência dos tópicos descritos relato do resumo, conforme o recomendado pelo checklist PRISMA for Abstrac	72
ANEXOS	74

ARTIGO - Versão em português***Propriedades psicométricas de questionários de frequência alimentar usados na odontologia: uma revisão sistemática***

Caroline Souza dos Santos ^a, Samuel Jorge Moysés ^b, Rodrigo Nunes Rached ^b, Marcia Helena Baldani Pinto ^c, Renata Iani Werneck ^b, Gil Guilherme Gasparello ^a, Juliana Schaia Rocha ^b

^aMSc, Pontifícia Universidade Católica do Paraná – PUCPR, School of Life Sciences, Dentistry Department, Curitiba, PR, Brazil.

^bProfessor, Pontifícia Universidade Católica do Paraná – PUCPR, School of Life Sciences, Dentistry Department, Curitiba, PR, Brazil.

^cProfessor, Universidade Estadual de Ponta Grossa – UEPG, Dentistry Department, Ponta Grossa, PR, Brazil.

Título abreviado: Questionário de propriedades psicométricas de frequência alimentar utilizado em odontologia

Caries Research
(<https://www.karger.com/Journal/Home/224219>)

Submetido em 28 jun 2022

RESUMO

Para analisar as propriedades psicométricas de instrumentos/escalas de frequência alimentar utilizados em odontologia foram incluídos estudos sobre a construção e validação de instrumentos para avaliação do consumo alimentar em odontologia. Foram excluídos estudos de revisão e cartas aos editores. As bases de dados consultadas foram PubMed (07/08/2020), Scopus (27/08/2020), Web of Science (27/08/2020), Cumulative Index to Nursing and Allied Health Literature (CINAHL) e Dentistry and Oral Sciences Source (via EBSCO) (28/08/2020), LILACS e BBO (25/08/2020), Literatura Cinza: Proquest (01/10/2020), Banco de Teses Capes (01/10/2020), Biblioteca Digital Brasileira de Teses e Dissertações (01/01 a 10/2020), Google Scholar (02/10/2020) e anais da International Association for Dental Research (IADR) (10/10/2020). Os artigos foram selecionados inicialmente pela leitura dos títulos e resumos, seguida da leitura do texto completo para confirmação dos critérios de elegibilidade, com auxílio do software Rayyan. A avaliação do risco de viés foi realizada de acordo com os padrões baseados em CONsensus para a seleção de instrumentos de medição de saúde (COSMIN). Foram extraídos os seguintes dados: autor, país, idioma do instrumento, amostra, instrumento de avaliação, estrutura do instrumento, tipo de alimento, instrumentos comparados com a condição clínica, adaptação de instrumento existente e propriedades psicométricas avaliadas. Os processos de seleção, análise de risco de viés e extração de dados foram realizados por dois avaliadores independentes. Sete estudos foram identificados. Os instrumentos disponíveis na literatura estavam apenas em inglês, japonês e malaio. Apenas um estudo realizou a tradução e validação transcultural de um instrumento, enquanto os demais foram estudos de construção. Um estudo não avaliou as propriedades psicométricas. Quanto à avaliação da qualidade e classificação geral dos estudos pelo checklist do COSMIN, todos foram considerados “inadequados”, sendo a confiabilidade (teste-reteste) a propriedade psicométrica mais validada; apenas um estudo realizou a validação de todas as propriedades psicométricas medidas no COSMIN. Quanto à qualidade dos instrumentos apresentados, todos os estudos foram classificados como “inadequados” na avaliação geral. É evidente que são necessários avanços no processo de validação desses instrumentos com métodos

adequados e completos de validação de suas propriedades psicométricas para minimizar erros de mensuração e evitar estimativas incorretas.

INTRODUÇÃO

O consumo elevado e frequente de alimentos ou bebidas açucarados é considerado um fator de risco para desenvolvimento de doenças crônicas não transmissíveis (DCNT), como diabetes, hipertensão, cardiovasculares e cárie dentária, além de promover outros danos à saúde geral ¹. Em relação à cárie dentária, a *American Dental Association* menciona que o consumo alimentar é um fator determinante para uma boa condição de saúde bucal e têm influência direta para o surgimento da cárie dentária ².

A cárie dentária é a doença multifatorial de maior prevalência³. Fatores socioeconômicos, culturais, exposição ao flúor, consumo excessivo e frequentes de alimentos com alto teor de açúcares livres são os principais fatores associados ao desenvolvimento da doença ^{1; 4; 5}. A cariogenicidade de alimentos ou bebidas é determinada pela presença e frequência de consumo de carboidratos fermentáveis de texturas pegajosas e amidos altamente processados, os quais são usados como substrato por microrganismos da cavidade bucal e, conseqüentemente, contribuem para a formação da lesão cariosa ⁴.

Determinar a cariogenicidade dos alimentos consumidos é fundamental para a classificação de risco de adoecimento do paciente. Isso exige do profissional da Odontologia uma avaliação precisa do controle dietético, com a escolha de um instrumento de avaliação dietética confiável, além de conhecimento, treinamento, e tempo para uma avaliação com qualidade ^{6; 7}. Na prática clínica, os instrumentos mais comuns para avaliar a ingestão alimentar são o recordatório de 24 horas, diário alimentar de três dias e os questionários de frequência alimentar (QFA). Os QFA são os mais frequentemente utilizados em estudos que investigam a associação entre o consumo dietético e a ocorrência de desfechos clínicos, porém, não está relatado na literatura qual a qualidade dos instrumentos disponíveis para uso na Odontologia ^{4; 7; 8}.

Para a escolha de um instrumento de avaliação dietética, principalmente para aplicação em nível populacional, devem ser eliminados ao máximo os possíveis erros de aferição devido às fontes de vieses sistemáticos, pelo processo de validação ⁹. Esse processo de validação é de suma importância para que os profissionais da Odontologia tenham disponível um método eficaz, fidedigno e reconhecido para diagnóstico de comportamentos alimentares de risco à cárie

dentária. São necessários instrumentos validados, considerando sua qualidade, aplicabilidade e suas propriedades psicométricas (consistência interna; confiabilidade teste-reteste; validade de conteúdo, validade de face, validade de construto), com destaque para os níveis de confiabilidade, validade, responsividade e interpretabilidade ¹⁰. Assim, os dados podem ser generalizados de maneira apropriada, ou seja, sabendo que a informação coletada mensura e avalia o consumo alimentar de forma confiável, medindo exatamente aquilo que se propõe medir ⁷.

Portanto, o objetivo desta revisão sistemática é analisar a qualidade dos instrumentos/escalas de frequência alimentar usados na Odontologia, considerando suas propriedades psicométricas.

MÉTODOS

Esta revisão teve seu protocolo submetido para registro na base PROSPERO (*International Prospective Register of Systematic Reviews*), sob protocolo número **CRD42021223949**, e realizada de acordo com as diretrizes do PRISMA 2020 (*Transparent Reporting of Systematic Reviews and Meta-Analyses*)¹¹.

A construção da pergunta do estudo seguiu o acrônimo CITYPOMP proposto por Munn et al. (2018)¹² para estudos de validação. A pergunta do estudo foi: Qual é a qualidade (confiabilidade, validade e capacidade de resposta) dos questionários de frequência alimentar para avaliar o consumo de alimentos com alto potencial cariogênico usados na Odontologia? O referido acrônimo orientou a seleção das palavras-chave:

Construct of interest - **Constructo de interesse (CI)**: Avaliar o consumo de alimentos com alto potencial cariogênico.

Type of measurement instrument - **Tipos de instrumento (TY)**: Questionário de Frequência alimentar.

Population - **População (PO)**: Pessoas atendidas em cuidados odontológicos.

Measurement properties - **Propriedades de mensuração (MP)**: confiabilidade, validade e capacidade de resposta.

As bases de dados usadas para o levantamento bibliográfico foram: Pubmed, Scopus, Web of Science, *Cumulative Index to Nursing and Allied Health Literature* (CINAHL) e Dentistry & Oral Sciences Source (via EBSCO) e *Latin American and Caribbean Health Sciences Literature database* – (Lilacs) e Bibliografia Brasileira de Odontologia (BBO). Foram empregados vocabulário controlado e termos livres relacionados por operadores booleanos “OR” e “AND”, de acordo com a questão norteadora do estudo. Cada estratégia de busca foi adaptada para as especificidades de cada base. Não houve restrição de idioma ou período de publicação. As estratégias de busca para todas as bases de dados estão incluídas na Tabela 1. Todos os resultados foram importados para o gerenciador de referências Endnote para remoção de duplicatas em seguida para o gerenciador Rayyan QCRI para realizar o processo de seleção.

A literatura cinzenta também foi acessada, para encontrar trabalhos não publicados nas bases principais: Proquest, Banco de Teses Capes e Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Google acadêmico e nos anais da *International Association for Dental Research* (IADR). A lista de referências dos artigos primários incluídos também foi verificada, a fim de identificar algum título não recuperado na busca principal. No período de 08/07/2020 a 10/10/2020 foram realizadas as buscas em todas as bases de dados e literatura cinzenta.

Tabela 1. Estratégia de busca de acordo com cada base de dados. Data das buscas de 08/07/2020 a 10/10/2020

#1 AND #2 AND #3 – PUBMED (08.07.2020)	
#	(((((("Dental caries"[MeSH Terms]) OR ("Dental caries"[Title/Abstract])) OR ("caries"[Title/Abstract])) OR ("Streptococcus mutans"[MeSH Terms])) OR ("Streptococcus mutans"[Title/Abstract])) OR (Cariogenic[Title/Abstract])) OR (Cariogenicity[Title/Abstract])) OR ("Dental Decay"[Title/Abstract])) OR ("Tooth Decay"[Title/Abstract])) OR (decay[Title/Abstract])) OR (carious[Title/Abstract])) OR ("Dental caries risk"[Title/Abstract])) OR ("Dentistry"[MeSH Terms])) OR ("Dentistry"[Title/Abstract]))
#	(((((("Nutrition Assessment"[MeSH Terms]) OR ("Nutrition Assessment"[Title/Abstract])) OR ("Dietary questionnaire"[Title/Abstract])) OR ("Food Frequency Questionnaire"[Title/Abstract])) OR ("food assessment"[Title/Abstract])) OR ("Health Impact Assessment"[MeSH Terms])) OR ("24h Recalls"[Title/Abstract])) OR ("Diet evaluation"[Title/Abstract])) OR ("Feeding Behavior"[MeSH Terms])) OR ("Feeding Behavior"[Title/Abstract])) OR ("Consumption analysis"[Title/Abstract])) OR ("Diet Surveys"[MeSH Terms])) OR ("Diet Surveys"[Title/Abstract])) OR (Diet[MeSH Terms])) OR (Diet[Title/Abstract])) OR ("Food questionnaire"[Title/Abstract])) OR ("Food frequency"[Title/Abstract])) OR ("Food consumption"[Title/Abstract])) OR ("food diary"[Title/Abstract])) OR ("Sugar ingestion"[Title/Abstract])) OR ("Free sugars"[Title/Abstract]))
#	(((((("Validation Study" [Publication Type]) OR ("Validation study"[Title/Abstract])) OR ("Validity of results"[Title/Abstract])) OR ("Reproducibility of Results"[MeSH Terms])) OR (Validity[Title/Abstract])) OR ("Validation process"[Title/Abstract])) OR ("Validating"[Title/Abstract])) OR (Reproducibility[Title/Abstract])) OR (Reliable[Title/Abstract])) OR ("Sensitivity and Specificity"[MeSH Terms])) OR (Sensitivity[Title/Abstract])) OR (Specificity[Title/Abstract])) OR (Unreliable[Title/Abstract])) OR (Reliable[Title/Abstract])) OR (Questionnaires[Title/Abstract])) OR (feasibility[Title/Abstract])) OR (instrument[Title/Abstract])) OR (instruments[Title/Abstract])) OR (measure[Title/Abstract])) OR (measures[Title/Abstract])) OR (validation[Title/Abstract])) OR (interpretability[Title/Abstract])) OR (Responsiveness[Title/Abstract]))
#1 AND #2 AND #3 - WEB OF SCIENCE (27.08.2020)	
#1	TÓPICO: ("Validation Study") OR TÓPICO: (Reproducibility) OR TÓPICO: (Validation process) OR TÓPICO: (Reliable*) OR TÓPICO: (Sensitivity) OR TÓPICO: (Specificity) OR TÓPICO: (Questionnaire\$) OR TÓPICO: (Instrument\$) OR TÓPICO: (Interpretability) OR TÓPICO: (Responsiveness) OR TÓPICO: (Feasibility)
#2	TÓPICO: ("Nutrition Assessment") OR TÓPICO: (Dietary questionnaire\$) OR TÓPICO: ("Food Frequency Questionnaire") OR TÓPICO: ("Food assessment") OR TÓPICO: ("Health Impact Assessment") OR TÓPICO: ("24h Recalls") OR TÓPICO: ("Diet evaluation") OR TÓPICO: ("Feeding Behavior") OR TÓPICO: ("Consumption analysis") OR TÓPICO: ("Diet Survey\$") OR TÓPICO: ("Food consumption") OR TÓPICO: ("food diary") OR TÓPICO: ("Sugar ingestion") OR TÓPICO: ("Free sugars")
#3	TÓPICO: ("Dental caries") OR TÓPICO: (Carie*) OR TÓPICO: ("Streptococcus mutans") OR TÓPICO: (mutans) OR TÓPICO: (Cariogenic*) OR TÓPICO: ("Dental Decay") OR TÓPICO: ("Tooth Decay") OR TÓPICO: (Decay) OR TÓPICO: (Carious*) OR TÓPICO: ("Dental caries risk") OR TÓPICO: (Decay*) OR TÓPICO: ("Dental caries risk") OR TÓPICO: (Dentist*)

#1 AND #2 AND #3 – SCOPUS (27.08.2020)	
# 1	((TITLE-ABS-KEY ("Validation Study") OR TITLE-ABS-KEY (reproducibility) OR TITLE-ABS-KEY (valid*) OR TITLE-ABS-KEY ("Validation process") OR TITLE-ABS-KEY (reliable) OR TITLE-ABS-KEY (sensitivity) OR TITLE-ABS-KEY (specificity) OR TITLE-ABS-KEY (unreliable) OR TITLE-ABS-KEY (questionnaire) OR TITLE-ABS-KEY (feasibility) OR TITLE-ABS-KEY (instrument) OR TITLE-ABS-KEY (measure) OR TITLE-ABS-KEY (responsiveness) OR TITLE-ABS-KEY (interpretability))))
# 2	((TITLE-ABS-KEY ("Nutrition Assessment") OR TITLE-ABS-KEY ("Dietary questionnaire") OR TITLE-ABS-KEY ("Food Frequency Questionnaire") OR TITLE-ABS-KEY ("Food assessment") OR TITLE-ABS-KEY ("Health Impact Assessment") OR TITLE-ABS-KEY ("24h Recalls") OR TITLE-ABS-KEY ("Diet evaluation") OR TITLE-ABS-KEY ("Feeding Behavior") OR TITLE-ABS-KEY ("Consumption analysis") OR TITLE-ABS-KEY ("Diet Surveys") OR TITLE-ABS-KEY ("Diet Food questionnaire") OR TITLE-ABS-KEY (diet*) OR TITLE-ABS-KEY ("Food consumption") OR TITLE-ABS-KEY ("food diary") OR TITLE-ABS-KEY ("Food frequency") OR TITLE-ABS-KEY ("Sugar ingestion") OR TITLE-ABS-KEY ("Free sugars"))))
# 3	((TITLE-ABS-KEY ("Validation Study") OR TITLE-ABS-KEY (reproducibility) OR TITLE-ABS-KEY (valid*) OR TITLE-ABS-KEY ("Validation process") OR TITLE-ABS-KEY (reliable) OR TITLE-ABS-KEY (sensitivity) OR TITLE-ABS-KEY (specificity) OR TITLE-ABS-KEY (unreliable) OR TITLE-ABS-KEY (questionnaire) OR TITLE-ABS-KEY (feasibility) OR TITLE-ABS-KEY (instrument) OR TITLE-ABS-KEY (measure) OR TITLE-ABS-KEY (responsiveness) OR TITLE-ABS-KEY (interpretability))))
#1 AND #2 AND #3 - PORTAL REGIONAL DA BVS (LILACS e BBO) (25.08.2020)	
# 1	(tw:((tw:("Dental caries")) OR (tw:("Cáries dentárias")) OR (tw:("Caries dental")) OR (tw:(Caries)) OR (tw:(cárie)) OR (tw:(Streptococcus mutans)) OR (tw:(Cariogenic)) OR (tw:(cariogênico)) OR (tw:(cariogénico)) OR (tw:(Cariogenicidade)) OR (tw:(Cariogenidad)) OR (tw:(Dental Decay)) OR (tw:(Cárie dentária)) OR (tw:(Cavidad dental)) OR (tw:(Tooth Decay)) OR (tw:(carie dentária)) OR (tw:(Caries dental)) OR (tw:(Decay)) OR (tw:(Cariouss)) OR (tw:(cariado)) OR (tw:(decaído)) OR (tw:(Dental caries risk)) OR (tw:(risco de cárie dentária)) OR (tw:(riesgo de caries dental)) OR (tw:(Dentistry)) OR (tw:(odontologia)) OR (tw:(odontología))))
# 2	(tw:((tw:(Nutrition Assessment)) OR (tw:(Avaliação Nutricional)) OR (tw:(evaluación nutricional)) OR (tw:(Dietary questionnaire)) OR (tw:(questionário dietético,)) OR (tw:(Cuestionario dietético)) OR (tw:(Food Frequency Questionnaire)) OR (tw:(questionário de frequência alimentar)) OR (tw:(cuestionario de frecuencia alimentaria)) OR (tw:(food assessment)) OR (tw:(avaliação alimentar)) OR (tw:(evaluación de alimentos)) OR (tw:(Health Impact Assessment)) OR (tw:(avaliação de impacto na saúde)) OR (tw:(evaluación de impacto en la salud)) OR (tw:(24h Recalls)) OR (tw:(recordatório 24horas)) OR (tw:(recuerdos 24h)) OR (tw:(Diet evaluation)) OR (tw:(avaliação de dieta)) OR (tw:(evaluación de la dieta)) OR (tw:(Feeding Behavior)) OR (tw:(comportamento alimentar)) OR (tw:(comportamiento de alimentación)) OR (tw:(Consumption analysis)) OR (tw:(análise de consumo)) OR (tw:(análisis de consumo)) OR (tw:(Diet Surveys)) OR (tw:(pesquisa de dieta)) OR (tw:(encuestas dietéticas)) OR (tw:(Diet)) OR (tw:(dieta)) OR (tw:(Food questionnaire)) OR (tw:(questionário alimentar)) OR (tw:(cuestionario de alimentos)) OR (tw:(Food frequency)) OR (tw:(frecuencia alimentar)) OR (tw:(frecuencia alimentaria)) OR (tw:(Food consumption)) OR (tw:(consumo

alimentar)) OR (tw:(food diary)) OR (tw:(diário alimentar)) OR (tw:(diário de comida))
OR (tw:(Sugar ingestion)) OR (tw:(ingestão de açúcar)) OR (tw:(ingestión de
azúcar)) OR (tw:(Free sugars)) OR (tw:(açúcares livres)) OR (tw:(azúcares libres)))

3 (tw:((tw:(Validation)) OR (tw:(validação)) OR (tw:(validación)) OR (tw:(Validation
Study)) OR (tw:(estudo de validação)) OR (tw:(estudio de validación)) OR
(tw:(Validity of results)) OR (tw:(validade dos resultados)) OR (tw:(validez de los
resultados)) OR (tw:(Reproducibility of Results)) OR (tw:(reprodutibilidade dos
resultados)) OR (tw:(reproducibility of resultados)) OR (tw:(Validity)) OR
(tw:(validade)) OR (tw:(validez)) OR (tw:(Validation process)) OR (tw:(processo de
validação)) OR (tw:(proceso de validacion)) OR (tw:(Validating)) OR (tw:(validando))
OR (tw:(Reproducibility)) OR (tw:(reprodutibilidade)) OR (tw:(reproducibilidad)) OR
(tw:(Reliable)) OR (tw:(confiável)) OR (tw:(De confianza)) OR (tw:(Sensitivity and
Specificity)) OR (tw:(sensibilidade e especificidade)) OR (tw:(sensibilidad y
especificidad)) OR (tw:(Sensitivity)) OR (tw:(sensibilidade)) OR (tw:(sensibilidad))
OR (tw:(Specificity)) OR (tw:(especificidade)) OR (tw:(especificidad)) OR
(tw:(Unreliable)) OR (tw:(não confiável)) OR (tw:(No fidedigno)) OR
(tw:(Questionnaires)) OR (tw:(questionários)) OR (tw:(cuestionarios)) OR
(tw:(Feasibility)) OR (tw:(viabilidade)) OR (tw:(factibilidad)) OR (tw:(Instrument)) OR
(tw:(instrumento)) OR (tw:(Instruments)) OR (tw:(instrumentos)) OR (tw:(Measure))
OR (tw:(a medida)) OR (tw:(medida)) OR (tw:(Interpretability)) OR
(tw:(interpretabilidade)) OR (tw:(interpretabilidad)) OR (tw:(Responsiveness)) OR
(tw:(capacidade de resposta,)) OR (tw:(sensibilidad))))

#1 AND #2 AND #3 - CINAHL e Dentistry & Oral Sciences Source (via EBSCO)
(28.08.2020)

#1 (((TX "Dental caries" OR AB "Dental caries" OR SU "Dental caries" OR TX Carie\$
OR AB Carie\$ OR TX "Streptococcus mutans" OR AB "Streptococcus mutans" OR
SU "Streptococcus mutans" OR TX Cariogenic\$ OR AB Cariogenic\$ OR TX
Cariogenicity OR AB Cariogenicity) OR TX "Dental Decay" OR AB "Dental Decay"
OR TX "Tooth Decay" OR AB "Tooth Decay" OR TX Decay OR AB Decay OR TX
Carious OR AB Carious OR TX "Dental caries risk" OR AB "Dental caries risk") OR
TX Dentistry OR AB Dentistry OR SU Dentistry)

#2 ((((TX "Nutrition Assessment" OR AB "Nutrition Assessment" OR SU "Nutrition
Assessment" OR TX "Dietary questionnaire" OR AB "Dietary questionnaire" OR TX
"Food Frequency Questionnaire" OR AB "Food Frequency Questionnaire" OR TX
"food assessment" OR AB "food assessment" OR TX "Health Impact Assessment"
OR AB "Health Impact Assessment" OR SU "Health Impact Assessment") OR TX
"24h Recalls" OR AB "24h Recalls" OR TX "Diet evaluation" OR AB "Diet evaluation"
OR TX "Feeding Behavior" OR AB "Feeding Behavior" OR SU "Feeding Behavior"
OR TX "Consumption analysis" OR AB "Consumption analysis") OR TX Diet\$ OR
AB Diet\$ OR TX "Food questionnaire" OR AB "Food questionnaire" OR TX "Food
frequency" OR AB "Food frequency" OR TX "Food consumption" OR AB "Food
consumption" OR TX "food diary" OR AB "food diary") OR TX "Sugar ingestion" OR
AB "Sugar ingestion" OR TX "Sugar ingestion" OR AB "Sugar ingestion") OR TX
Free sugar\$ OR AB Free sugar\$)

#3 ((((TX Valid\$ OR AB Valid\$ OR TX "Validation Study" OR AB "Validation Study"
OR TX "Validity of result\$" OR AB "Validity of result\$" OR TX "Reproducibility of
Results" OR AB "Reproducibility of Results" OR SU "Reproducibility of Results" OR
TX Validity OR AB Validity) OR TX "Validation process" OR AB "Validation process"
OR TX Validating OR AB Validating OR TX "Reproducibility" OR AB

“Reproducibility” OR TX “Reliable” OR AB “Reliable” OR TX (“Sensitivity and Specificity”) OR AB (“Sensitivity and Specificity”) OR SU (“Sensitivity and Specificity”)) OR TX Sensitivity OR AB Sensitivity OR TX Specificity OR AB Specificity OR TX Unreliable OR AB Unreliable OR TX Questionnaire\$ OR SU Questionnaire\$ OR TX Feasibility OR AB Feasibility) OR TX Feasibility OR AB Feasibility OR TX Instrument\$ OR AB Instrument\$ OR TX Measure OR AB Measure OR TX Interpretability OR AB Interpretability OR TX Responsiveness OR AB Responsiveness)

Critério de Elegibilidade

Foram incluídos estudos de construção e/ou validação de instrumentos/escalas de avaliação do consumo alimentar para uso na Odontologia, com ou sem suas propriedades psicométricas avaliadas. Foram excluídos estudos de revisão, carta aos editores e protocolos de pesquisa.

Seleção dos Estudos

A seleção dos artigos incluídos foi inicialmente feita pela leitura dos títulos e resumos, sendo que duplicatas oriundas de ambas as bases foram eliminadas. Caso houvesse alguma dúvida na elegibilidade do estudo, por falta de informações no resumo, foi realizada a leitura completa. Após, todos os estudos incluídos foram lidos na íntegra para confirmação dos critérios de elegibilidade.

Todo o processo foi realizado por dois examinadores (C.S e S.V.R) de forma independente, obedecendo rigorosamente os critérios de inclusão e exclusão. As discordâncias foram resolvidas por meio de discussões, com o auxílio de um terceiro revisor (J.S.R).

Foi avaliada a concordância entre os revisores pelo coeficiente de Kappa, sendo considerado aceitável os valores acima de 0,7 a 0,8 substancial e 0,8 a 1,0 de concordância excelente ¹³. O valor encontrado após a leitura de títulos e resumos foi de 0,77, considerada uma concordância substancial. Este valor aumentou para 0,81 após leitura do texto completo, considerada uma concordância excelente ¹³.

Extração dos dados

Dois revisores realizaram a extração dos dados de forma independente (C.S e S.V.R), usando uma ficha de extração elaborada para esta revisão, de acordo com as características identificadas nos estudos. Antes da extração, a ficha foi pré-

testada e os pesquisadores foram treinados. As divergências entre os pesquisadores foram resolvidas por meio de discussão, com o auxílio de um terceiro revisor (J.S.R).

Avaliação de qualidade

A qualidade metodológica dos estudos incluídos foi avaliada pelo *Consensus-based Standards for the Selection of Health Measurement Instruments* (COSMIN)¹⁴, instrumento específico para estudos de validação. Este instrumento permite a avaliação da qualidade de estudos, considerando as seguintes propriedades psicométricas: consistência interna; confiabilidade teste-reteste; validade de conteúdo, validade de face, validade de construto¹⁰. Ele possui 10 categorias de avaliação, usadas para determinar se o estudo satisfaz as normas de boa qualidade, teoria de resposta e generalização dos resultados¹⁵.

As categorias 1 e 2 referem-se à validade do conteúdo (Requisitos, abrangência gerais do projeto e validação por especialistas); 3 - 5 à validade estrutural, consistência interna (CI) e validade transcultural, respectivamente; juntas, formam a estrutura interna. As categorias 6 - 10 formam as propriedades das medições restantes: confiabilidade, erro de medição, teste de hipótese para validade de construto e capacidade de resposta¹⁶. As categorias consistem em itens que avaliam a propriedade da medida. Cada item pode ser classificado em: (1) muito bom; (2) adequado; (3) duvidoso; ou (4) inadequado (5) não se aplica (NA). A pontuação final de cada categoria é determinada pela opção mais baixa dos itens classificado¹⁵.

A avaliação do risco de viés foi realizada de forma independente por dois examinadores (C.S e S.V.R). As divergências foram discutidas até se obter consenso, com a participação de um terceiro revisor (J.S.R.). Todos os pesquisadores foram previamente treinados para a utilização do COSMIN. A concordância entre os avaliadores foi calculada por meio do Coeficiente de Correlação Intraclassa (CCI)¹⁰. Também foi calculado para cada propriedade, por estudo, o percentual de itens classificados como "ótimo", "muito bom", "adequado", "duvidoso" ou "inadequado". Para isso, foi realizado o teste não-paramétrico de Wilcoxon para variável ordinal para avaliar se não ocorreu erro sistemático, sendo que o nível de significância adotado foi de $p \geq 0,05$, não podendo haver diferença

significante entre o escore médio da classificação de cada um dos examinadores. A concordância entre os examinadores (CCI) foi de 0,92, indicando uma concordância excelente. O teste não-paramétrico de Wilcoxon teve como resultado um valor de $p = 0,166$, portanto não houve erro sistemático entre os avaliadores (Apêndice 1).

Análise e síntese dos dados

Foi realizada análise de dados, considerando os seguintes itens: autor, país, idioma do instrumento, amostra, instrumento de avaliação, estrutura do instrumento, tipo de alimentos, comparação com condição clínica, adaptação de instrumento existente, finalidade do estudo e propriedades psicométricas avaliadas.

Não foi possível realizar a metanálise das propriedades psicométricas, pois apenas os estudos de Psoter, et al.¹⁷ e Devenish, et al.¹⁸ expressaram dados de correlação de Pearson e as estruturas destes instrumentos eram muito discrepantes, um deles é um questionário de frequência alimentar com avaliação de alimentos e bebidas com teor de açúcares totais, livres e protetores¹⁸. O outro é um questionário de frequência alimentar relativo do consumo de bebidas e doces açucarados e de avaliação da quantidade de açúcares adicionada no preparo dos alimentos¹⁷.

RESULTADOS

O processo de seleção dos artigos está descrito na figura 1. Após realizar as buscas nas bases de dados e literatura cinzenta foram encontradas um total de 6382 artigos, após a remoção de duplicatas. Foram selecionados 8 artigos após a leitura de títulos e resumos, porém na etapa de leitura na íntegra dos textos, um dos estudos foi excluído pois não era um instrumento específico para avaliação do consumo alimentar e do risco de cárie. Ao final, foram incluídos 7 artigos nesta revisão de literatura.

A tabela 2 traz um resumo das informações encontradas nos artigos incluídos ^{17; 19; 20; 21; 22; 23; 24}. A maioria dos estudos foi realizado no continente americano, nos países Estados Unidos ²⁴, Canadá ¹⁹ e Haiti ¹⁷. Foram realizados dois estudos no continente asiático, Japão ²¹ e Malásia²⁰ e dois no continente da Oceania, ambos na Austrália ^{1; 18}. Os instrumentos disponíveis na literatura estão apenas nos idiomas inglês ^{1; 6; 18; 24}, japonês ²¹, crioulo^{17;} e malaio ²⁰. Apenas um estudo realizou tradução e validação transcultural de um instrumento ²¹, os demais eram estudos de construção. Shinga-Ishihara, et al. ²¹ construiu e adaptou o instrumento de uma versão inglesa²⁵ para o idioma japonês. O estudo na sua versão inglesa ²⁵ não foi encontrada para leitura completa, impedindo de avaliar a qualidade e a validação do instrumento original, considerando suas propriedades psicométricas.

A validação completa do instrumento, considerando as categorias avaliadas pelo COSMIN, foi feita apenas pelo estudo de Shinga-Ishihara et al. ²¹. Foram identificados 5 estudos ^{1; 6; 17; 20; 24} que realizaram a análise de pelo menos uma propriedade psicométrica (validação linguística, confiabilidade e reprodutibilidade teste-reteste, painel de especialistas, teste piloto, validade de critério e consistência interna). Apenas um estudo não avaliou nenhuma de suas propriedades psicométricas ¹⁸.

Em relação à população-alvo dos instrumentos, os propostos por Amezdroz et al. ²² e Custodio-Lumsden et al. ²⁴ eram destinados à crianças, porém respondidos pelos pais, pois avaliavam a dieta de crianças até a idade pré-escolar. ^{22; 24}. O instrumento construído por Psoter et al. ¹⁷ realizou a avaliação dietética conjunta de pais e seus filhos. Quatro estudos validaram seus instrumentos para uma população adulta, sendo que Shinga- Ishihara et al. ²¹ aplicaram a versão

japonesa do QFA em gestantes ²¹. Os estudos de Ngah, et al. ²⁰ e Devenish et al. ¹⁸ obtiveram informações da frequência do consumo alimentar de pais de crianças, para avaliação do consumo habitual familiar. Apenas Patenaude et al. ⁶ validaram seu instrumento para uma população em geral.

Em relação aos objetivos dos instrumentos, alguns avaliaram a frequência alimentar do consumo geral, outros eram focados em alimentos com alto teor de açúcar ou associados a outras medidas. Estes com mais de uma medida avaliaram, além da frequência alimentar, o comportamento familiar ou outros fatores de risco relacionados ao desenvolvimento da cárie dentária. Devenish, et al. ¹⁸ avaliaram o consumo alimentar geral, considerando não só alimentos cariogênicos, mas também os alimentos protetores da saúde geral. Três estudos ^{17; 18; 21} avaliaram a frequência alimentar relativa apenas aos alimentos açucarados. Dois estudos elaboraram um QFA associado a outro método ^{6; 24}. O instrumento proposto por Patenaude, et al. ⁶ avaliou a frequência do consumo alimentar em conjunto com o comportamento alimentar. Custodio-Lumsden et al. ²⁴ desenvolveram uma ferramenta chamada *MySmileBuddy* (MSB), que avaliou, além do consumo alimentar, as práticas alimentares, atitudes e crenças do cuidador, uso do flúor e história familiar. Por fim, o instrumento de Amezdroz et al. ¹ era uma escala cariogênica, medindo a cariogenicidade de alimentos e bebidas.

Apenas os estudos de Shinga-Ishihara et al. ²¹, Custodio-Lumsden et al. ²⁴ e Amezdroz et al. ¹ realizaram a comparação dos resultados do instrumento proposto com a condição clínica do paciente. Esses estudos realizaram uma comparação da frequência do consumo alimentar com os parâmetros clínicos: presença de cárie, utilizando exames clínicos bucais (ICDAS II) ¹ e avaliação do nível de *streptococos mutans* salivares ^{21; 24}.

A tabela 3 apresenta a classificação da qualidade dos estudos para cada uma das propriedades psicométricas consideradas no Checklist COSMIN ¹⁴. Em relação a avaliação geral dos estudos, todos foram classificados como “inadequados” ^{1;6;17;18; 20; 21; 24}. No geral, dentre todas as categorias avaliadas pelo COSMIN, a categoria estrutura interna - validade estrutural (3) foi a mais frequentemente avaliada como adequada. Já a categoria validade de construto (9)

foi a pior avaliada, tendo em vista que ela não foi considerada pela maioria dos estudos e apenas

foram classificados como “adequado” ^{21; 24}. O estudo Shinga-Ishihara et al. ²¹ recebeu a melhor avaliação, visto que foi o único estudo a avaliar todas as propriedades psicométricas previstas no COSMIN, porém nem todos os critérios de cada categoria foram cumpridos e por isso ele também teve sua classificação geral como “inadequado”.

Todos os estudos avaliaram as categorias confiabilidade (Teste reteste)^{1;6; 17;20;24}. Porém, apenas um estudo ²¹ obteve classificação de qualidade “adequado” na categoria confiabilidade (Teste reteste). Para avaliação da capacidade de resposta, os estudos de Shinga-Ishihara, et al. ²¹, Devenish. et al. ²³ e Custodio-Lumsden, et al.²⁶ obtiveram a classificação de “duvidoso”. Os demais apresentaram qualidade “inadequada” para esta categoria^{1; 6; 17; 20}, visto que nenhum artigo validou esta propriedade psicométrica de forma completa.

Figura 1. Fluxograma do processo de seleção do estudo de acordo com as diretrizes do PRISMA 2020.

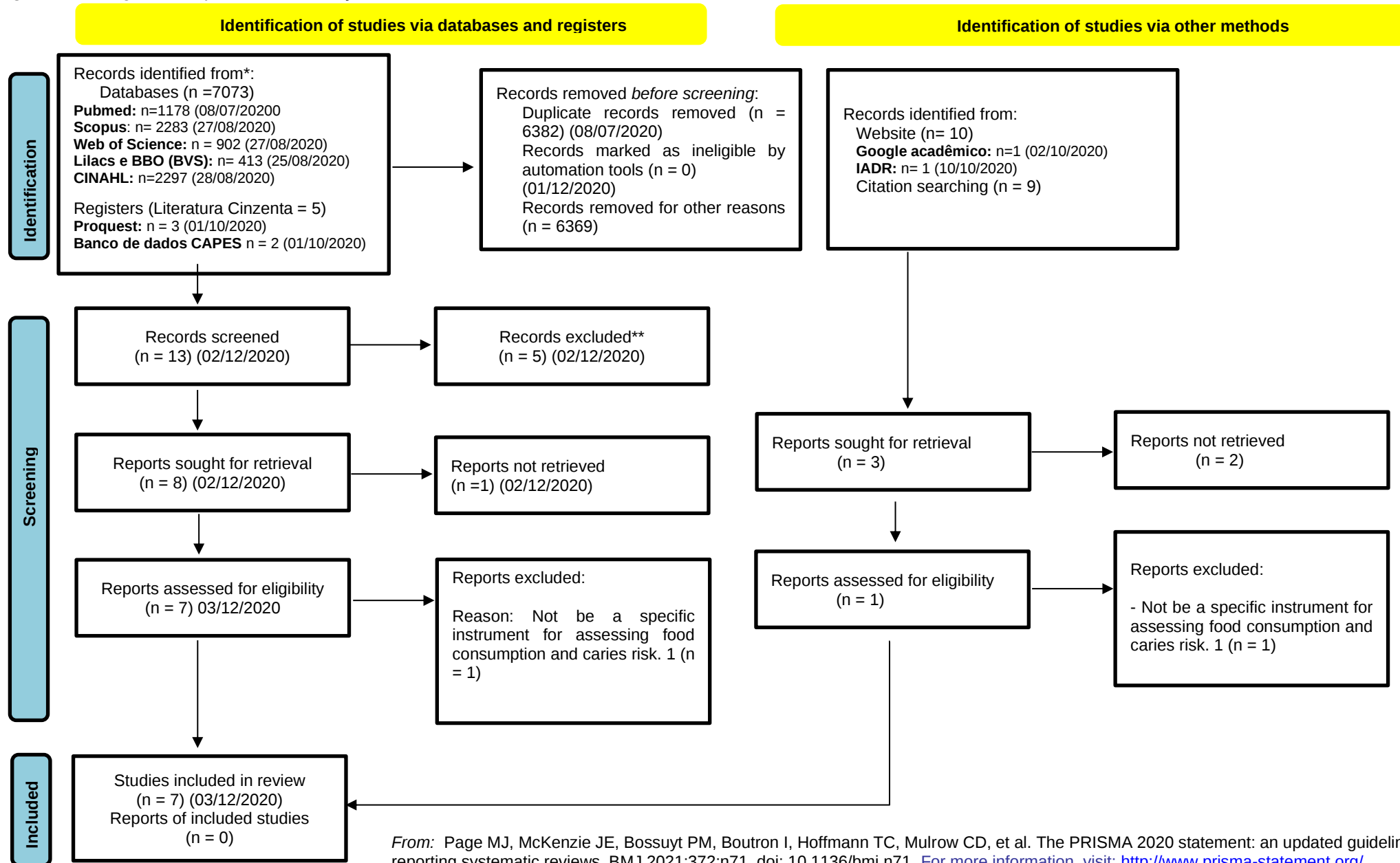


Tabela 2. Resumo dos estudos incluídos nesta revisão sistemática.

Autor	País	Idioma do instrumento	Amostra	Instrumento de avaliação	Estrutura do instrumento	Tipo de alimentos	Comparação com condição clínica	Adaptação de instrumento existente	Finalidade do estudo	Propriedades psicométricas avaliadas
Amezdroz, et al,2019 ¹	Austrália	Inglês	Criança (entrevista com os pais) 6 meses n= 377 12 meses n= 330 18 meses n= 387 5 anos n = 259	Consumo usual das crianças: QFA/Escala Cariogênica	32 itens alimentares / 8 respostas/ Composição Nutricional	Alimentos e bebidas de acordo com o grau de cariogenicidade	Exames Clínicos Orais (ICDAS II)	Não	Construção e Validação	Validade de constructo; pré-teste; Reprodutibilidade (teste reteste)
Patenaude, et al,2020 ⁶	Canadá	Inglês	População em geral Feminino n= 67 Masculino n= 23	QFA/ Comportamento Alimentar	7 blocos de frequência alimentar/ 23 itens alimentares	Alimentos com carboidratos fermentáveis e com ação direta de açúcar	Não	Não	Construção e validação	Especialistas; teste piloto; Confiabilidade (teste reteste)
Psoter, et al, 2008 ¹⁷	Haiti	Crioulo	Mãe e filho (9-17 anos) Mães n=30 Filhos n=30	QFA do consumo de açúcar e de adição de açúcar no preparo dos alimentos	5questões/filho (consumo de açúcar) 5questões/mãe (quantidade de açúcar adicionado)	Bebidas e doces Açucarados	Não	Sim	Validação	Confiabilidade (teste reteste), pré-teste, validade de conteúdo, tradução
Ngah, et al,2019 ²⁰	Malásia	Malaio	Pais de crianças (6 a 11 anos) Feminino n= 40 Masculino n= 9	QFA Cariogênica e saúde bucal	Versão Inglesa traduzida para Bahasa Melayu 15 categorias de alimentos e bebidas/26 itens alimentares	Alimentos e bebidas industrializados cariogênicos	Não	Sim	Validação Linguística	Validação Linguística; Confiabilidade (teste reteste); teste piloto, especialistas; conteúdo

Autor	País	Idioma do instrumento	Amostra	Instrumento de avaliação	Estrutura do instrumento	Tipo de alimentos	Comparação com condição clínica	Adaptação de instrumento existente	Finalidade do estudo	Propriedades psicométricas avaliadas
Shinga-Ishihara, et al, 2014 ²¹	Japão	Japonês	Gestantes (19 a 43 anos) n=355	QFA Açucarados e industrializados	QFA versão traduzida para o Japonês - 38 itens alimentares.	Alimentos e bebidas relacionados à cárie	Nível estreptococos mutans salivares Dentocult SM Strip Mutans (Orion Diagnostica, Espoo, Finlândia)	Sim	Validação Transcultural	Validação de constructo; consistência Interna; Confiabilidade (teste-reteste) e Intraexaminador, critério; traducultural; teste piloto; conteúdo
Custodio-Lumsden et al, 2015 ²⁴	EUA	Inglês	Crianças (2 a 6 anos) Feminino n= 96 Masculino n= 12	<i>MySmileBuddy</i> (MSB)	NA	NA	Nível estreptococos mutans salivares (amostra de saliva)	Não	Validação	Confiabilidade (teste-reteste) Intraexaminador; validade de critério e constructo
Devenish, et al ¹⁸	Austrália	Inglês	Pais de crianças (18 a 36 meses) n= 97	QFA	QFA 24 blocos/89 itens alimentares/4 perguntas descritivas	Alimentos e bebidas com teor de açúcar total, livre e protetores	Não	Não	Construção	NA

NA= Não aplicável

Tabela 3. Classificação da qualidade dos estudos para cada uma das propriedades psicométricas de acordo com o Checklist COSMIN.

Autor/Ano	Categoria 1	Categoria 2	Categoria 3	Categoria 4	Categoria 5	Categoria 6	Categoria 7	Categoria 8	Categoria 9	Categoria 10	Final
Amezdroz et al. 2019 ¹	Inadequado	NA	NA	NA	NA	Inadequado	Inadequado	Inadequado	Inadequado	Inadequado	Inadequado
Patenaude et al. 2020 ⁶	Inadequado	Inadequado	Duvidoso	NA	NA	Inadequado	Inadequado	Inadequado	NA	Inadequado	Inadequado
Psoter et al. 2008 ¹⁷	Inadequado	Inadequado	Adequado	Inadequado	Duvidoso	Duvidoso	Inadequado	Inadequado	NA	Inadequado	Inadequado
Ngah et al. 2019 ²⁰	Inadequado	Duvidoso	Adequado	Inadequado	Adequado	Inadequado	Duvidoso	Inadequado	NA	Inadequado	Inadequado
Shinga-Ishihara et al. 2014 ²¹	Inadequado	Duvidoso	Adequado	Adequado	Adequado	Adequado	Adequado	Inadequado	Adequado	Duvidoso	Inadequado
Custodio-Lumsden 2016 ²⁴	NA	Inadequado	Inadequado	Inadequado	NA	Inadequado	Inadequado	Inadequado	Adequado	Duvidoso	Inadequado
Devenish, et al. 2019 ¹⁸	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Inadequado

NA= Não aplicável.

Categoria 1: Requisitos gerais (Entrevista cognitiva/teste piloto)

Categoria 2: Validade de conteúdo (Especialistas)

Categoria 3: Estrutura interna: Validade estrutural

Categoria 4: Estrutura interna: Consistência interna

Categoria 5: Estrutura interna: Validade transcultural

Categoria 6: Confiabilidade (Teste reteste)

Categoria 7: Erro de medição

Categoria 8: Teste de hipótese

Categoria 9: Validade de construto

Categoria 10: Capacidade de resposta

Final: Classificação final de cada categoria, determinada pela opção mais baixa.

DISCUSSÃO

A avaliação do consumo alimentar é uma parte determinante na abordagem do tratamento odontológico ⁷, pois a ingestão, em especial de alimentos e bebidas açucarados, está diretamente relacionada com o desenvolvimento da cárie dentária ^{14; 27}. A correta avaliação deste consumo representa um grande desafio para estes profissionais, sendo de suma importância o uso de instrumentos confiáveis, com suas propriedades psicométricas validadas ¹⁰ e capazes de identificar alimentos com risco potencial para doenças bucais de maneira eficaz ^{4;7}.

A maioria dos instrumentos encontrados nesta revisão são questionários de frequência alimentar relativos ao consumo de alimentos e bebidas com alto teor de açúcares totais e livres, direcionados para a faixa etária de crianças até a idade pré-escolar ^{1; 17; 18; 20; 26} e em inglês ^{1; 6; 17; 18; 24}. Em relação à qualidade desses instrumentos, todos os estudos foram classificados como “inadequados” na avaliação geral, sendo que a confiabilidade (teste-reteste) foi a propriedade psicométrica mais validada ^{1; 6; 17; 18; 20; 21} e apenas um estudo ²¹ realizou a validação de todas as propriedades psicométricas previstas no COSMIN, mesmo que de forma incompleta.

A avaliação do consumo alimentar por meio de Questionários de Frequência Alimentar (QFA) permite que pesquisas epidemiológicas e clínicas sejam conduzidas para avaliar o papel da dieta na etiologia de diversas enfermidades, como no surgimento da cárie dentária ^{4; 7}. Dentre os estudos incluídos, é predominante o desenvolvimento e validação de questionários de frequência alimentar relativos ao consumo de açúcares totais em alimentos e bebidas ^{1; 6; 17; 20}. Esses instrumentos permitem a investigação e avaliação direta da regularidade com que o paciente tem acesso a determinados alimentos com alto potencial cariogênico e a associação de tal consumo com desfechos clínicos. Eles também podem ser usados para avaliação individual do paciente na clínica odontológica, pois permitem a personalização do aconselhamento dietético, bem como estimulam a discussão entre profissional e paciente ^{4; 7; 28; 29}.

Um achado deste estudo que chama a atenção é a predominância de instrumentos no idioma em inglês, sendo que apenas três eram em outras línguas, japonesa, crioulo e malaio ^{17;20; 21}. Porém a avaliação de dieta é realizada em estudos da área da Odontologia de diversos países e idiomas ^{1; 20; 30; 31; 32}, pois ela

é um dos fatores relacionados à cárie dentária. Em relação aos métodos de avaliação de dieta mais frequentemente usados em estudos publicados na literatura, duas revisões sistemáticas,^{31; 32; 33} encontraram que a maioria dos estudos utilizaram o método do Diário Alimentar (DA), seguido do Questionário de Frequência Alimentar (QFA), para avaliar a relação entre cárie dentária e o consumo alimentar. O principal achado desses estudos foi que o consumo em alta frequência de alimentos açucarados associou-se a uma maior experiência de cárie dentária, estudos esses prospectivos e com populações de crianças e adolescentes ^{31; 32}.

Questionário de Frequência Alimentar (QFA), Diário Alimentar (DA) e Recordatório 24horas (R24h) são os métodos de referência a serem utilizados em estudos epidemiológicos e/ou na prática clínica para avaliação do consumo alimentar ^{4; 7; 34}. Para a escolha do melhor método, deve-se considerar o objetivo da análise (individual ou epidemiológica), as propriedades psicométricas e a possibilidade de realizar as adaptações da medida para a população de estudo, considerando as diversidades regionais sobre os hábitos alimentares ^{21; 35}. Tanto o DA e/ou R24h podem ser aplicados em qualquer população, mas para estudos epidemiológicos estes instrumentos não englobam variabilidade, hábitos e a sazonalidade da dieta, e pouco representam o consumo habitual ⁹. Além disso, a compilação e análise de seus resultados é mais complexa pela natureza qualitativa dos seus dados, sendo de difícil aplicação em grandes estudos epidemiológicos^{7; 8}. Já o QFA faz a estimativa do consumo dietético ao longo do tempo, baseada na possibilidade de medir a intensidade da exposição e classificar os indivíduos de acordo com a frequência do consumo, o que possibilita a estimativa da associação das categorias de consumo com o desenvolvimento de DCNT ^{4; 7}. Porém, pelo motivo desses instrumentos avaliarem o consumo alimentar a longo prazo, pode ocorrer viés de memória e subestimação das quantidades absolutas de consumo⁷.

Essa variabilidade de métodos encontrada na literatura pode dificultar a comparabilidade entre os estudos, pois, como visto, alguns não permitem uma avaliação em longo prazo, como o DA e o R24h. Considerando a complexidade na obtenção de medidas válidas da ingestão alimentar em estudos epidemiológicos, reforça-se a necessidade de adequação e validação de instrumentos para essa

finalidade^{7; 35}. Além disso a formulação de instrumentos em conjunto com profissionais da Nutrição exerce papel importante na construção e avaliação do consumo alimentar de forma confiável.

A maioria dos QFA disponíveis foi construída e validada para avaliar a frequência do consumo alimentar de alimentos e bebidas açucaradas e o seu grau de cariogenicidade em crianças até a idade pré-escolar^{1; 17; 18; 20; 26}. Avaliar o consumo alimentar nesta idade é relevante na Odontologia, pois a cárie dentária pode evoluir rapidamente, sendo que qualquer sinal de lesão antes dos três anos já é considerado cárie severa³⁶. O alto consumo de açúcar nos primeiros anos de vida pode ser explicado por diversos aspectos, os quais ocorrem simultaneamente³⁷. As expectativas dos pais em relação à quantidade e frequência alimentar de seus filhos podem estar em desacordo com as necessidades fisiológicas das crianças, principalmente pelo possível incentivo ao aumento no consumo de alimentos com a oferta de doces, visto que nesta fase há maior resistência para o consumo de alimentos. Deve-se também considerar que hábitos alimentares obedecem a padrão cultural e simbólico e a oferta de produtos cariogênicos pode ter o significado de afeto e amor.^{4; 7; 37}.

Ainda na faixa etária em crianças até a idade pré-escolar, as escolhas alimentares e frequência da alimentação são totalmente dependentes dos pais e cuidadores;³⁸ além disso, as crianças não têm habilidade cognitiva para autorrelatarem o seu consumo alimentar de forma completa¹⁸. Sendo assim, esses instrumentos devem ser respondidos por pais e cuidadores^{1; 18; 20}. Há limitações para obtenção de informações sobre alimentação da criança por meio dos pais e cuidadores, visto que em sua maioria não podem permanecer com os filhos grande parte do dia e informações podem ser ocultadas.^{35; 39}. Porém estes instrumentos não deixam de ser relevantes, tendo em vista a importância de se identificar fatores de risco à cárie, incluindo a dieta.^{4; 7} Isso reforça a importância da avaliação da validade de constructo desses instrumentos, principalmente a validade convergente, comparando seus resultados com condições clínicas do paciente. Na literatura, são poucos os instrumentos validados para adultos e idosos⁴⁰, e, uma vez que a carga da doença atinge a idade adulta, é importante que a relação entre dieta e cárie dentária seja investigada também nesses grupos^{31; 32; 40}.

O estudo de Shinga-Ishihara et al.²¹ foi o único que fez a tradução e adaptação transcultural de um instrumento já existente em inglês para o japonês, elaborado para ser aplicado em crianças²⁵. A versão original possui 35 itens de alimentos, com avaliação do consumo alimentar relacionado à cárie durante o mês anterior ²⁵, porém este instrumento não foi encontrado na sua versão completa, para que suas propriedades psicométricas fossem avaliadas. A versão japonesa²¹ adicionou alimentos de consumo regular no país para aumentar a validade do conteúdo. Nesta versão, o questionário foi validado exclusivamente para a população de gestantes, visto que preferências dietéticas frequentemente mudam neste período e podem levar à ingestão frequente de alimentos cariogênicos ²¹. Os estudos de Psoter, et al.¹⁷ e Ngah et al.²⁰ realizaram apenas a tradução do instrumento, sendo traduzido/retrotraduzido de uma língua de origem para outra língua utilizando falantes nativos, não foi relatado se foi realizado a adaptação transcultural na versão teste-reteste. O estudo de Psoter, et al.¹⁷, realizou apenas a validação da reprodutibilidade (teste-reteste). Não foi encontrado na sua versão original, para que a construção e suas propriedades psicométricas fossem avaliadas por falta de acesso das referências citadas no instrumento.

Para validação transcultural, adaptações devem ser feitas de forma criteriosa, visto que o instrumento deve ser confiável internamente e para comparações internacionais, garantindo a qualidade do dado e da informação produzida. O instrumento deve ser traduzido/retrotraduzido de uma língua de origem para outra língua utilizando falantes nativos e com domínio na outra língua, realizado por pares de forma oculta, realizando a comparabilidade dos resultados de tradução, adaptação da cultura, hábitos, sazonalidade da dieta e a realização de pré-teste para modificações necessárias²¹, que demonstre confiabilidade e validade ^{20; 21}. No estudo de Shinga-Ishihara et al.²¹, foram realizadas todas as etapas de tradução e adaptação transcultural, garantindo a confiabilidade e a comparabilidade com o estudo original ²⁰, porém houve falhas no processo de validação, os critérios de cada categoria não foram cumpridos de forma completa e por isso ele também teve sua classificação geral como “inadequado”.

Três estudos ^{1;24;21} realizaram a validação de constructo, com a comparação dos resultados do instrumento proposto com a condição clínica do paciente. Nos

achados do estudo de Custodio-Lumsden et al.²⁴ e Shinga-Ishihara et al.²¹ os níveis de estreptococos mutans salivares apresentaram uma relação estatisticamente significativa com consumo dietético, indicando que níveis mais elevados estão associados a maiores pontuações do consumo alimentar cariogênico. O estudo de Amezdroz, et al.¹ com a utilização dos Exames Clínicos Bucais (ICDAS II) apresentou um aumento significativo de lesão de cárie como doença avançada aos cinco anos da criança, uma vez que ela inicia uma ingestão muito maior de alimentos discricionários, alimentos que não fornecem nutrientes, incluindo, dieta cariogênica. Essas relações do instrumento com a condição clínica são fundamentais para que conhecimentos sejam generalizados de maneira efetiva, e a informação reflita a condição real do indivíduo e possar medir exatamente aquilo que se propõe⁷.

A propriedade mais frequentemente avaliada nos estudos foi a confiabilidade (teste-reteste)^{1; 6; 17; 20; 21; 26}. Nela, utiliza-se de repetições de medidas para avaliar a confiabilidade do instrumento, a qual evidencia o nível de correlação existente entre as distribuições de escores num mesmo teste em duas ocasiões diferentes e independentes^{41; 42; 43}. Se a correlação entre os resultados das duas aplicações for significativamente positiva, o instrumento pode ser considerado confiável⁴². Apesar de muito avaliada, apenas um estudo²¹ obteve classificação de qualidade “adequado” de acordo com o COSMIN, visto que os estudos^{1; 6; 17; 20; 21; 24} validaram essa propriedade de forma incompleta, com a ausência do tempo de intervalo entre duas aplicações, falta da descrição do treinamento e se os entrevistadores eram aptos a realizar a pesquisa, colocando em dúvida a qualidade do instrumento de medida^{41; 42}. Além disso, apesar de ser uma qualidade necessária, ela deve ser avaliada em conjunto com outros componentes, dentro das propriedades psicométricas buscando garantir a sua precisão^{10; 42}.

O check-list COSMIN é uma ferramenta que visa a excelência na qualidade metodológica de estudos de propriedade de medidas¹⁴. Os resultados gerais encontrados mostram que todos os estudos incluídos nessa revisão foram classificados como “inadequados”, visto que a classificação geral da qualidade de cada estudo é determinada sempre pela pior condição encontrada nos critérios avaliados em cada categoria. O estudo de Shinga-Ishihara et al.²¹ recebeu a

melhor avaliação, de acordo com a escala COSMIN, pois foi o único a avaliar todas as propriedades psicométricas previstas. Porém ele ainda foi classificado como inadequado, pois as categorias de teste de hipótese e capacidade de respostas estavam incompletas²¹. Isso mostra que a maioria dos estudos apresentam limitações importantes que influenciam a evidência sobre os instrumentos, se realmente medem aquilo que se propõe medir⁴⁴. A rigidez da classificação geral do COSMIN se dá pelo fato de que, para que um instrumento tenha validade, confiabilidade e reprodutibilidade, deve-se considerar cuidadosamente a psicometria de cada medida^{10; 15; 42}.

É possível destacar outras limitações que afetaram a qualidade metodológica do estudo incluídos nesta revisão, tais como a limitação da população da amostra, falha no detalhamento de todos as propriedades psicométricas realizadas^{6; 17; 20; 21; 24}. Dentre as principais falhas, um estudo¹⁸ não realizou nenhuma de suas propriedades psicométricas para validação do estudo. Quando uma psicometria do instrumento não é testada, sua medida será imprecisa, o que torna limitada a validade das inferências obtidas a partir desse instrumento ⁴¹. É necessário realizar comparação desses instrumentos com as condições clínicas de saúde bucal, para garantir que estão medindo exatamente o que se espera medir ^{7; 42}.

Em relação às limitações desta revisão sistemática, é importante ressaltar que não foi possível ter acesso na íntegra de dois estudos ^{21;25}, o que pode ter limitado a cobertura dessa revisão. Porém, foram feitas todas as tentativas possíveis para ter acesso aos artigos completos.

CONCLUSÃO

A maioria dos instrumentos encontrados nesta revisão de literatura foi composta de questionários de frequência alimentar relativos ao consumo de alimentos e bebidas com alto teor de açúcares totais e livres, direcionados para a faixa etária de crianças até a idade pré-escolar. Todos os estudos foram classificados como “inadequados” na avaliação geral, a confiabilidade (teste-reteste) foi a propriedade psicométrica mais validada e apenas um estudo realizou a validação de todas as propriedades psicométricas previstas no COSMIN. São necessários avanços no processo de validação desses instrumentos, com métodos apropriados e completos para validação de suas propriedades psicométricas, para minimizar erros de medição e evitar estimativas incorretas.

REFERÊNCIAS

- 1 AMEZDROZ, E. et al. Feasibility and development of a cariogenic diet scale for epidemiological research. **International Journal of Paediatric Dentistry**, Malden, Massachusetts, v. 29, n. 3, p. 310-324, 2019. ISSN 0960-7439. Disponível em: <
[http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=136497029](http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=136497029&)
& >.
- 2 ASSOCIATION, A. D. **American Dental Association** 2022. Disponível em: <
<https://www.ada.org/resources/research/science-and-research-institute/oral-health-topics/nutrition-and-oral-health>>.
- 3 MOREIRA, P. R. **Práticas alimentares relacionadas à cárie dentária : uma revisão**. Universidade Federal do Rio Grande do Sul: UFRGS, 2016. 42 Disponível em: < <http://hdl.handle.net/10183/151466> >.
- 4 ROSA, Q. F. D. **Development and validation of a Food Frequency Questionnaire to assess the adolescent's consumption of foods with cariogenic and erosive potential**. UFPel. Brasil: Universidade Federal de Pelotas, 2015. 74 Disponível em: <
<http://repositorio.ufpel.edu.br:8080/handle/prefix/3535> >.
- 5 SAÚDE, M. D. **Saúde Bucal**. 2008. Disponível em: <
https://bvsmis.saude.gov.br/bvs/publicacoes/saude_bucal.pdf >.
- 6 PATENAUDE, S. A.; PAPAGERAKIS, P.; LIEFFERS, J. R. L. Development of a nutrition questionnaire for dental caries risk factors. **International Journal of Environmental Research and Public Health**, v. 17, n. 5, 2020. Disponível em: <
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081538785&doi=10.3390%2fijerph17051793&partnerID=40&md5=6a5e75569853c320f4fb80dc98678635> >.
- 7 FERNANDES, M. P. **Construção e validação de um Questionário de Frequência Alimentar para avaliar o consumo de alimentos cariogênicos em pré-escolares**. 2015. Disponível em: <
<http://guaiaca.ufpel.edu.br/handle/prefix/3892>>.
- 8 PEDRAZA, D. F.; MENEZES, T. N. [Food Frequency Questionnaire developed and validated for the Brazilian population: a review of the literature]. **Cien Saude Colet**, v. 20, n. 9, p. 2697-720, Sep 2015. ISSN 1413-8123.
- 9 FISBERG, R.; MARCHIONI, D.; COLUCCI, A. Avaliação do consumo alimentar e da ingestão de nutrientes na prática clínica. **Arquivos Brasileiros De Endocrinologia E Metabologia - ARQ BRAS ENDOCRINOL METABOL**, v. 53, 07/01 2009.
- 10 CAMPOS, M. et al. Propriedades psicométricas de questionários de atividade física na adolescência: revisão sistemática. **Revista Brasileira de Atividade Física & Saúde**, v. 23, p. 1-13, 03/26 2019.

- 11 PAGE, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **Bmj**, v. 372, p. n71, Mar 29 2021. ISSN 0959-8138 (Print) 0959-8138.
- 12 MUNN, Z. et al. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. **BMC Medical Research Methodology**, v. 18, n. 1, p. 5, 2018/01/10 2018. ISSN 1471-2288. Disponível em: < <https://doi.org/10.1186/s12874-017-0468-4> >.
- 13 LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-74, Mar 1977. ISSN 0006-341X (Print) 0006-341x.
- 14 MOKKINK, L. B. et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. **Qual Life Res**, v. 27, n. 5, p. 1171-1179, May 2018. ISSN 0962-9343 (Print) 0962-9343.
- 15 PRUDÊNCIO, D. A. et al. Questionnaires and scales for assessment of ankle function: a systematic review of instruments translated and validated for Brazilian Portuguese. **Disability and Rehabilitation**, v. 43, n. 3, p. 309-316, 2021/01/30 2021. ISSN 0963-8288. Disponível em: < <https://doi.org/10.1080/09638288.2019.1626917> >.
- 16 MELO, B. M. et al. **A importância da prevenção baseada no IHOS**, 2014. p. Disponível em: < <http://revistas.faculdefacit.edu.br/index.php/JOFI/article/view/35> >
- 17 PSOTER, W. J.; GEBRIAN, B.; KATZ, R. V. Reliability of a sugar consumption questionnaire for rural Haiti. **Puerto Rico Health Sciences Journal**, v. 27, n. 1, p. 69-74, 2008. Disponível em: < <https://www.scopus.com/inward/record.uri?eid=2-s2.0-47049111214&partnerID=40&md5=24f8a6418e1bc5b1c76041dba9e88dde> >.
- 18 DEVENISH, G. et al. Development and Relative Validity of a Food Frequency Questionnaire to Assess Intakes of Total and Free Sugars in Australian Toddlers. **Int J Environ Res Public Health**, v. 14, n. 11, Nov 8 2017. ISSN 1661-7827 (Print) 1660-4601.
- 19 S, A. P.; PAPAGERAKIS, P.; J, R. L. L. Development of a Nutrition Questionnaire for Dental Caries Risk Factors. **Int J Environ Res Public Health**, v. 17, n. 5, Mar 10 2020. ISSN 1661-7827 (Print) 1660-4601.
- 20 NGAH, R.; HASAN, R.; RAHMAN, N. A. Linguistic validity and reliability of cariogenic food frequency and oral health knowledge, attitude and practice questionnaires for parents of 6-11 years old children in Kota Bharu, Kelantan. **Malaysian Journal of Nutrition**, v. 25, n. 1, p. 79-97, 2019. ISSN 1394-035X. Disponível em: <

[http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=138293156
& > .](http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=138293156& > .)

- 21 SHINGA-ISHIHARA, C. et al. Cross-cultural validity of a dietary questionnaire for studies of dental caries risk in Japanese. **BMC Oral Health**, v. 14, p. 1, Jan 2 2014. ISSN 1472-6831.
- 22 AMEZDROZ, E. et al. Feasibility and development of a cariogenic diet scale for epidemiological research. **Int J Paediatr Dent**, v. 29, n. 3, p. 310-324, May 2019. ISSN 0960-7439.
- 23 DEVENISH, G. et al. Early childhood feeding practices and dental caries among Australian preschoolers. **American Journal of Clinical Nutrition**, v. 111, n. 4, p. 821-828, 2020. Disponível em: <
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084239207&doi=10.1093%2fajcn%2fnqaa012&partnerID=40&md5=1d790b96170e3983c7333ee2cbeb7a66> >.
- 24 CUSTODIO-LUMSDEN, C. L. et al. Validation of an early childhood caries risk assessment tool in a low-income Hispanic population. **Journal of Public Health Dentistry**, Malden, Massachusetts, v. 76, n. 2, p. 136-142, Spring 2016 2016. ISSN 0022-4006. Disponível em: <
[http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=115831087
& > .](http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=115831087& > .)
- 25 EVENS , C. C. **Snacking patterns as a risk factor for early childhood caries..** 1997. University de Washington. UMI number 9736267 Disponível em: <
https://alliance-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=TN_cdi_proquest_journals_304390795&context=PC&vid=UW&lang=en_US&search_scope=all&adaptor=primo_central_multiple_fe&tab=default_tab&query=any,contains,Evens%20CC:%20Snacking%20patterns%20as%20a%20risk%20factor%20for%20early%20childhood%20caries.%20PhD>
- 26 CUSTODIO-LUMSDEN, C. L. et al. Validation of an early childhood caries risk assessment tool in a low-income Hispanic population. **J Public Health Dent**, v. 76, n. 2, p. 136-42, Mar 2016. ISSN 0022-4006 (Print) 0022-4006.
- 27 TINÍ, G. F.; LONG, S. M. **Avaliação de diários alimentares de crianças atendidas na clínica infantil de uma universidade privada de São Paulo**, 2015. p. Disponível em: <
<https://www.metodista.br/revistas/revistas-metodista/index.php/Odonto/article/view/6979>>
- 28 ARHEIAM, A. et al. Are diet diaries of value in recording dietary intake of sugars? A retrospective analysis of completion rates and information quality. **Br Dent J**, v. 221, n. 9, p. 571-576, Nov 4 2016. ISSN 0007-0610.
- 29 ARHEIAM, A. et al. Reasons for low adherence to diet-diaries issued to pediatric dental patients: a collective case study. **Patient Prefer Adherence**, v. 12, p. 1401-1411, 2018. ISSN 1177-889X (Print) 1177-889x.

- 30 ARHEIAM, A. et al. The use of diet diaries in general dental practice in England. **Community Dent Health**, v. 33, n. 4, p. 267-273, Dec 2016. ISSN 0265-539X (Print) 0265-539x.
- 31 HANCOCK, S.; ZINN, C.; SCHOFIELD, G. The consumption of processed sugar- and starch-containing foods, and dental caries: a systematic review. **European Journal of Oral Sciences**, v. 128, n. 6, p. 467-475, 2020. ISSN 0909-8836. Disponível em: <
<https://onlinelibrary.wiley.com/doi/abs/10.1111/eos.12743> >.
- 32 MAHBOOBI, Z. et al. Dietary free sugar and dental caries in children: A systematic review on longitudinal studies. **Health Promot Perspect**, v. 11, n. 3, p. 271-280, 2021. ISSN 2228-6497 (Print) 2228-6497.
- 33 MOYNIHAN, P. J.; KELLY, S. A. M., ISSN 1. Effect on Caries of Restricting Sugars Intake: Systematic Review to Inform **WHO Guidelines** ISSN 1. Disponível em:<
<https://journals.sagepub.com/doi/abs/10.1177/0022034513508954>>
- 34 FERNANDES, M. L. D. M. F. et al. Cárie dentária e necessidade de tratamento ortodôntico: impacto na qualidade de vida de escolares. **Pesqui. bras. odontopediatria clín. integr**, v. 13, n. 1, 2013/03 2013. Disponível em: <
<http://revista.uepb.edu.br/index.php/pboci/article/view/1379/965> >.
- 35 KOLODZIEJCZYK, J. K.; MERCHANT, G.; NORMAN, G. J. Reliability and validity of child/adolescent food frequency questionnaires that assess foods and/or food groups. **J Pediatr Gastroenterol Nutr**, v. 55, n. 1, p. 4-13, Jul 2012. ISSN 0277-2116.
- 36 NOWAK, A. J. Oral health policies and clinical guidelines. **Pediatr Dent**, v. 29, n. 2, p. 138-9, Mar-Apr 2007. ISSN 0164-1263 (Print) 0164-1263.
- 37 YEE, A. Z.; LWIN, M. O.; HO, S. S. The influence of parental practices on child promotive and preventive food consumption behaviors: a systematic review and meta-analysis. **Int J Behav Nutr Phys Act**, v. 14, n. 1, p. 47, Apr 11 2017. ISSN 1479-5868.
- 38 SCAGLIONI, S. et al., ISSN 6. Factors Influencing Children's Eating Behaviours ISSN 6. Disponível em:
<<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6024598/>>
- 39 MORIKAVA, F. S. et al. Healthy and cariogenic foods consumption and dental caries: A preschool-based cross-sectional study. **Oral Diseases**, Malden, Massachusetts, v. 24, n. 7, p. 1310-1317, 2018. ISSN 1354-523X. Disponível em: <
<http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=131754754>
& >.

- 40 MOYNIHAN, P. J.; KELLY, S. A. Effect on caries of restricting sugars intake: systematic review to inform WHO guidelines. **J Dent Res**, v. 93, n. 1, p. 8-18, Jan 2014. ISSN 0022-0345 (Print) 0022-0345.
- 41 ALVES, I. B. **Validade e confiabilidade do questionário nórdico de sintomas musculoesqueléticos: uma revisão sistemática de literatura**. 2017. Universidade Federal da Bahia Faculdade de Medicina da Bahia Mestrado em Saúde, Ambiente e Trabalho, Universidade Federal da Bahia Faculdade de Medicina da Bahia Disponível em:<
<https://repositorio.ufba.br/ri/bitstream/ri/31269/1/Dissertacao%20Final%20-%20Ivone.pdf>>
- 42 CUNHA, C.; PEREIRA DE ALMEIDA NETO, O.; STACKFLETH, R. Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. **Revista Brasileira Ciências da Saúde - USCS**, v. 14, 08/18 2016.
- 43 ALEXANDRE, N. et al., et al., **Propriedades psicométricas na avaliação de instrumentos: avaliação da confiabilidade e da validade**. Disponível em:<
<https://www.scielo.br/j/ress/a/v5hs6c54VrhmjvN7yGcYb7b/abstract/?lang=pt>>
- 44 VERMELHO, N. F. F. **Propriedades de medida dos instrumentos de avaliação do equilíbrio e risco de quedas na doença de Parkinson: uma revisão sistemática**. 2014. Medicina de reabilitação, Universidade Federal de Minas Gerais

MANUSCRIPT – English version***Psychometric properties of food frequency questionnaire used in dentistry: a systematic review***

Caroline Souza dos Santos ^a, Samuel Jorge Moysés ^b, Rodrigo Nunes Rached ^b, Marcia Helena Baldani Pinto ^c, Renata Iani Werneck ^b, Gil Guilherme Gasparello ^a, Juliana Schaia Rocha ^b

^aMSc, Pontifical Catholic University of Paraná, School of Life Sciences, Dentistry Department, Curitiba, PR, Brazil.

^bTeacher, Pontifical Catholic University of Paraná, School of Life Sciences, Dentistry Department, Curitiba, PR, Brazil.

^cTeacher, State University of Ponta Grossa, Dentistry Department, Ponta Grossa, PR, Brazil.

Short Title: *Psychometric properties of food frequency questionnaire used in dentistry*

Caries Research
(<https://www.karger.com/Journal/Home/224219>)

Submitted in 28 jun 2022

ABSTRACT

To analyze the psychometric properties of food frequency instruments/scales used in dentistry studies on the construction and validation of instruments to assess food consumption in dentistry were included. Review studies and letters to the editors were excluded. The databases consulted were PubMed (07/08/2020), Scopus (08/27/2020), Web of Science (08/27/2020), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and Dentistry and Oral Sciences Source (via EBSCO) (28/08/2020), LILACS and BBO (25/08/2020), Gray Literature: Proquest (01/10/2020), Capes Theses Bank (01/10/2020), Brazilian Digital Library of Theses and Dissertations (01/01 10/2020), Google Scholar (02/10/2020), and proceedings of the International Association for Dental Research (IADR) (10/10/2020). Articles were initially selected by reading the titles and abstracts, followed by reading the full text to confirm the eligibility criteria, with the aid of Rayyan software. Risk of bias assessment was performed according to the CONsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. The following data were extracted: author, country, instrument language, sample, assessment instrument, instrument structure, type of food, instruments compared with the clinical condition, adaptation of an existing instrument, and psychometric properties evaluated. The selection, risk of bias analysis, and data extraction processes were performed by two independent evaluators. Seven studies were identified. The instruments available in the literature were only in English, Japanese, and Malay. Only one study performed translation and cross-cultural validation of an instrument, whereas the others were construction studies. One study did not assess psychometric properties. Regarding the quality assessment and general classification of the studies by the COSMIN checklist, all were considered “inadequate,” with reliability (test-retest) being the most validated psychometric property; only one study carried out the validation of all psychometric properties measured in COSMIN. Regarding the quality of the instruments presented, all the studies were classified as “inadequate” in the general assessment. It is evident that advances are needed in the validation process of these instruments with appropriate and complete methods for validating their psychometric properties to minimize measurement errors and avoid incorrect estimations.

INTRODUCTION

High and frequent consumption of sugary foods or drinks is considered a risk factor for the development of chronic non-communicable diseases, such as diabetes, hypertension, cardiovascular diseases, and dental caries, in addition to causing other damages to general health [Amezdroz, Carpenter et al. 2019]. The American Dental Association states that food consumption is a determining factor for good oral health and directly influences the development of dental caries [Association 2022].

Dental caries is the most prevalent multifactorial disease [Moreira 2016]. Socioeconomic and cultural factors, exposure to fluoride, and excessive and frequent consumption of foods with a high free-sugar content are the main factors associated with the development of the disease [Saúde 2008; Rosa 2015; Amezdroz, Carpenter et al. 2019]. The cariogenicity of foods or beverages is determined by the presence and frequency of consumption of fermentable carbohydrates with sticky textures and highly processed starches. These carbohydrates are used as a substrate by microorganisms in the oral cavity, and consequently contribute to the formation of carious lesions [Rosa 2015].

Determining the cariogenicity of the foods consumed is essential for classifying the patients' risk of illness. In addition to knowledge, training, and time for quality assessment, dental professionals require an accurate assessment of dietary control with the choice of a reliable dietary assessment instrument [Fernandes 2015; Patenaude, Papagerakis et al. 2020]. In clinical practice, the most common instruments to assess food intake are the 24-hour recall (R24h), three-day food diary (DA), and food frequency questionnaire (FFQ). FFQs are most frequently used in studies investigating the association between dietary consumption and clinical outcomes; however, the quality of the instruments available for use in dentistry has not been reported in the literature [Fernandes 2015; Pedraza and Menezes 2015; Rosa 2015].

When choosing a dietary assessment instrument, especially for application at the population level, possible errors in measurement due to sources of systematic bias should be eliminated by the validation process [Fisberg, Marchioni et al. 2009]. This validation process is of paramount importance for dental professionals in order to have an effective, reliable, and recognized method for

diagnosing eating behaviors at risk for dental caries. For the validation of the instruments, their quality, applicability, and psychometric properties (internal consistency, test-retest reliability, content validity, face validity, and construct validity), with emphasis on the levels of reliability, validity, responsiveness, and interpretability, need to be considered [Campos, Felicidade et al. 2019]. Thus, the data can be generalized in an appropriate way; that is, knowing that the information collected measures and evaluates food consumption reliably, measuring exactly what it intends to measure [Fernandes 2015]. Therefore, the objective of this systematic review was to analyze the quality of the food frequency instruments/scales used in dentistry while considering their psychometric properties.

METHODS

This review protocol was submitted for registration in the PROSPERO database (International Prospective Register of Systematic Reviews), under protocol number CRD42021223949, and performed according to the PRISMA 2020 guidelines (Transparent Reporting of Systematic Reviews and Meta-Analyses) [Page, McKenzie et al. 2021].

The construction of the study question followed the acronym CITYPOMP proposed [Munn, Stern et al. 2018] for the validation studies. The study question was as follows: What is the quality (reliability, validity, and responsiveness) of FFQs used in dentistry to assess the consumption of foods with high cariogenic potential? The acronym guided the selection of keywords.

Construct of interest (CI): To evaluate the consumption of foods with high cariogenic potential.

Type of measurement instrument (TY): FFQ.

Population (PO): People treated in dental care.

Measurement properties (MP): reliability, validity, and responsiveness.

The databases used for the bibliographic survey were PubMed, Scopus, Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Dentistry and Oral Sciences Source (via EBSCO), Latin American and Caribbean Health Sciences Literature Database (Lilacs), and Brazilian Bibliography of Dentistry (BBO). Controlled vocabulary and free terms related by Boolean operators “OR” and “AND” were used according to the guiding question of the study. Each search strategy was adapted according to the specifics of each database. There were no restrictions on language or publication period. The search strategies for all databases are included in Table 1. All results were imported into the endnote reference manager to remove duplicates and then into the Rayyan QCRI manager to perform the selection process.

Gray literature was also accessed from Proquest, Banco de Teses Capes and Biblioteca Digital de Teses e Dissertações (BDTD), Google Scholar, and the proceedings of the International Association for Dental Research (IADR) to find

unpublished works in the main databases. The reference lists of the primary articles included were also checked to identify any titles not retrieved in the main search. Searches were carried out in all databases and gray literature from July 08, 2020, to October 10, 2020.

Table 1. Search strategy according to each database. Date of searches from July 08th 2020 to October 10th 2020

#1 AND #2 AND #3 – PUBMED (08.07.2020)	
#1	((((((((((("Dental caries"[MeSH Terms]) OR ("Dental caries"[Title/Abstract])) OR ("caries"[Title/Abstract])) OR ("Streptococcus mutans"[MeSH Terms]) OR ("Streptococcus mutans"[Title/Abstract])) OR (Cariogenic[Title/Abstract])) OR (Cariogenicity[Title/Abstract])) OR ("Dental Decay"[Title/Abstract])) OR ("Tooth Decay"[Title/Abstract])) OR (decay[Title/Abstract])) OR (carious[Title/Abstract])) OR ("Dental caries risk"[Title/Abstract])) OR ("Dentistry"[MeSH Terms]) OR ("Dentistry"[Title/Abstract]))
#2	((((((((((("Nutrition Assessment"[MeSH Terms]) OR ("Nutrition Assessment"[Title/Abstract])) OR ("Dietary questionnaire"[Title/Abstract])) OR ("Food Frequency Questionnaire"[Title/Abstract])) OR ("food assessment"[Title/Abstract])) OR ("Health Impact Assessment"[MeSH Terms]) OR ("24h Recalls"[Title/Abstract])) OR ("Diet evaluation"[Title/Abstract])) OR ("Feeding Behavior"[MeSH Terms]) OR ("Feeding Behavior"[Title/Abstract])) OR ("Consumption analysis"[Title/Abstract])) OR ("Diet Surveys"[MeSH Terms]) OR ("Diet Surveys"[Title/Abstract])) OR (Diet[MeSH Terms]) OR (Diet[Title/Abstract])) OR ("Food questionnaire"[Title/Abstract])) OR ("Food frequency"[Title/Abstract])) OR ("Food consumption"[Title/Abstract])) OR ("food diary"[Title/Abstract])) OR ("Sugar ingestion"[Title/Abstract])) OR ("Free sugars"[Title/Abstract]))
#3	((((((((((("Validation Study" [Publication Type]) OR ("Validation study"[Title/Abstract])) OR ("Validity of results"[Title/Abstract])) OR ("Reproducibility of Results"[MeSH Terms]) OR (Validity[Title/Abstract])) OR ("Validation process"[Title/Abstract])) OR ("Validating"[Title/Abstract])) OR (Reproducibility[Title/Abstract])) OR (Reliable[Title/Abstract])) OR ("Sensitivity and Specificity"[MeSH Terms]) OR (Sensitivity[Title/Abstract])) OR (Specificity[Title/Abstract])) OR (Unreliable[Title/Abstract])) OR (Reliable[Title/Abstract])) OR (Questionnaires[Title/Abstract])) OR (feasibility[Title/Abstract])) OR (instrument[Title/Abstract])) OR (instruments[Title/Abstract])) OR (measure[Title/Abstract])) OR (measures[Title/Abstract])) OR (validation[Title/Abstract])) OR (interpretability[Title/Abstract])) OR (Responsiveness[Title/Abstract]))
#1 AND #2 AND #3 - WEB OF SCIENCE (27.08.2020)	
#1	TÓPICO: ("Validation Study") OR TÓPICO: (Reproducibility) OR TÓPICO: (Validation process) OR TÓPICO: (Reliable*) OR TÓPICO: (Sensitivity) OR TÓPICO: (Specificity) OR TÓPICO: (Questionnaire\$) OR TÓPICO: (Instrument\$) OR TÓPICO: (Interpretability) OR TÓPICO: (Responsiveness) OR TÓPICO: (Feasibility)
#2	TÓPICO: ("Nutrition Assessment") OR TÓPICO: (Dietary questionnaire\$) OR TÓPICO: ("Food Frequency Questionnaire") OR TÓPICO: ("Food assessment") OR TÓPICO: ("Health Impact Assessment") OR TÓPICO: ("24h Recalls") OR TÓPICO: ("Diet evaluation") OR TÓPICO: ("Feeding Behavior") OR TÓPICO: ("Consumption analysis") OR TÓPICO: ("Diet Survey\$") OR TÓPICO: ("Food consumption") OR TÓPICO: ("food diary") OR TÓPICO: ("Sugar ingestion") OR TÓPICO: ("Free sugars")
#3	TÓPICO: ("Dental caries") OR TÓPICO: (Carie*) OR TÓPICO: ("Streptococcus mutans") OR TÓPICO: (mutans) OR TÓPICO: (Cariogenic*) OR TÓPICO: ("Dental Decay") OR TÓPICO: ("Tooth Decay") OR TÓPICO: (Decay) OR TÓPICO: (Carious*)

OR TÓPICO: ("Dental caries risk") OR TÓPICO: (Decay*) OR TÓPICO: ("Dental caries risk") OR TÓPICO: (Dentist*)

#1 AND #2 AND #3 – SCOPUS (27.08.2020)

#1 ((TITLE-ABS-KEY ("Validation Study") OR TITLE-ABS-KEY (reproducibility) OR TITLE-ABS-KEY (valid*) OR TITLE-ABS-KEY ("Validation process") OR TITLE-ABS-KEY (reliable) OR TITLE-ABS-KEY (sensitivity) OR TITLE-ABS-KEY (specificity) OR TITLE-ABS-KEY (unreliable) OR TITLE-ABS-KEY (questionnaire) OR TITLE-ABS-KEY (feasibility) OR TITLE-ABS-KEY (instrument) OR TITLE-ABS-KEY (measure) OR TITLE-ABS-KEY (responsiveness) OR TITLE-ABS-KEY (interpretability))))

#2 ((TITLE-ABS-KEY ("Nutrition Assessment") OR TITLE-ABS-KEY ("Dietary questionnaire") OR TITLE-ABS-KEY ("Food Frequency Questionnaire") OR TITLE-ABS-KEY ("Food assessment") OR TITLE-ABS-KEY ("Health Impact Assessment") OR TITLE-ABS-KEY ("24h Recalls") OR TITLE-ABS-KEY ("Diet evaluation") OR TITLE-ABS-KEY ("Feeding Behavior") OR TITLE-ABS-KEY ("Consumption analysis") OR TITLE-ABS-KEY ("Diet Surveys") OR TITLE-ABS-KEY ("Diet Food questionnaire") OR TITLE-ABS-KEY (diet*) OR TITLE-ABS-KEY ("Food consumption") OR TITLE-ABS-KEY ("food diary") OR TITLE-ABS-KEY ("Food frequency") OR TITLE-ABS-KEY ("Sugar ingestion") OR TITLE-ABS-KEY ("Free sugars"))))

#3 ((TITLE-ABS-KEY ("Validation Study") OR TITLE-ABS-KEY (reproducibility) OR TITLE-ABS-KEY (valid*) OR TITLE-ABS-KEY ("Validation process") OR TITLE-ABS-KEY (reliable) OR TITLE-ABS-KEY (sensitivity) OR TITLE-ABS-KEY (specificity) OR TITLE-ABS-KEY (unreliable) OR TITLE-ABS-KEY (questionnaire) OR TITLE-ABS-KEY (feasibility) OR TITLE-ABS-KEY (instrument) OR TITLE-ABS-KEY (measure) OR TITLE-ABS-KEY (responsiveness) OR TITLE-ABS-KEY (interpretability))))

#1 AND #2 AND #3 - REGIONAL PORTAL OF BVS (LILACS e BBO) (25.08.2020)

#1 (tw:((tw:("Dental caries")) OR (tw:("Cáries dentárias")) OR (tw:("Caries dental")) OR (tw:(Caries)) OR (tw:(cárie)) OR (tw:(Streptococcus mutans)) OR (tw:(Cariogenic)) OR (tw:(cariogênico)) OR (tw:(cariogénico)) OR (tw:(Cariogenicidade)) OR (tw:(Cariogenidad)) OR (tw:(Dental Decay)) OR (tw:(Cárie dentária)) OR (tw:(Cavidad dental)) OR (tw:(Tooth Decay)) OR (tw:(carie dentária)) OR (tw:(Caries dental)) OR (tw:(Decay)) OR (tw:(Cariouss)) OR (tw:(cariado)) OR (tw:(decaído)) OR (tw:(Dental caries risk)) OR (tw:(risco de cárie dentária)) OR (tw:(riesgo de caries dental)) OR (tw:(Dentistry)) OR (tw:(odontologia)) OR (tw:(odontología))))

#2 (tw:((tw:(Nutrition Assessment)) OR (tw:(Avaliação Nutricional)) OR (tw:(evaluación nutricional)) OR (tw:(Dietary questionnaire)) OR (tw:(questionário dietético,)) OR (tw:(Cuestionario dietético)) OR (tw:(Food Frequency Questionnaire)) OR (tw:(questionário de frequência alimentar)) OR (tw:(cuestionario de frecuencia alimentaria)) OR (tw:(food assessment)) OR (tw:(avaliação alimentar)) OR (tw:(evaluación de alimentos)) OR (tw:(Health Impact Assessment)) OR (tw:(avaliação de impacto na saúde)) OR (tw:(evaluación de impacto en la salud)) OR (tw:(24h Recalls)) OR (tw:(recordatório 24horas)) OR (tw:(recuerdos 24h)) OR (tw:(Diet evaluation)) OR (tw:(avaliação de dieta)) OR (tw:(evaluación de la dieta)) OR (tw:(Feeding Behavior)) OR (tw:(comportamento alimentar)) OR (tw:(comportamiento de alimentación)) OR (tw:(Consumption analysis)) OR (tw:(análise de consumo)) OR (tw:(análisis de consumo)) OR (tw:(Diet Surveys)) OR (tw:(pesquisa de dieta)) OR (tw:(encuestas dietéticas)) OR (tw:(Diet)) OR (tw:(dieta)) OR (tw:(Food questionnaire))

OR (tw:(questionário alimentar)) OR (tw:(cuestionario de alimentos)) OR (tw:(Food frequency)) OR (tw:(frequencia alimentar)) OR (tw:(frecuencia alimentaria)) OR (tw:(Food consumption)) OR (tw:(consumo alimentar)) OR (tw:(food diary)) OR (tw:(diário alimentar)) OR (tw:(diário de comida)) OR (tw:(Sugar ingestion)) OR (tw:(ingestão de açúcar)) OR (tw:(ingestión de azúcar)) OR (tw:(Free sugars)) OR (tw:(açúcares livres)) OR (tw:(azúcares libres))))

#3

(tw:((tw:(Validation)) OR (tw:(validação)) OR (tw:(validación)) OR (tw:(Validation Study)) OR (tw:(estudo de validação)) OR (tw:(estudio de validación)) OR (tw:(Validity of results)) OR (tw:(validade dos resultados)) OR (tw:(validez de los resultados)) OR (tw:(Reproducibility of Results)) OR (tw:(reprodutibilidade dos resultados)) OR (tw:(reproducibility of resultados)) OR (tw:(Validity)) OR (tw:(validade)) OR (tw:(validez)) OR (tw:(Validation process)) OR (tw:(processo de validação)) OR (tw:(proceso de validacion)) OR (tw:(Validating)) OR (tw:(validando)) OR (tw:(Reproducibility)) OR (tw:(reprodutibilidade)) OR (tw:(reproducibilidad)) OR (tw:(Reliable)) OR (tw:(confiável)) OR (tw:(De confianza)) OR (tw:(Sensitivity and Specificity)) OR (tw:(sensibilidade e especificidade)) OR (tw:(sensibilidad y especificidad)) OR (tw:(Sensitivity)) OR (tw:(sensibilidade)) OR (tw:(sensibilidad)) OR (tw:(Specificity)) OR (tw:(especificidade)) OR (tw:(especificidad)) OR (tw:(Unreliable)) OR (tw:(não confiável)) OR (tw:(No fidedigno)) OR (tw:(Questionnaires)) OR (tw:(questionários)) OR (tw:(cuestionarios)) OR (tw:(Feasibility)) OR (tw:(viabilidade)) OR (tw:(factibilidad)) OR (tw:(Instrument)) OR (tw:(instrumento)) OR (tw:(Instruments)) OR (tw:(instrumentos)) OR (tw:(Measure)) OR (tw:(a medida)) OR (tw:(medida)) OR (tw:(Interpretability)) OR (tw:(interpretabilidade)) OR (tw:(interpretabilidad)) OR (tw:(Responsiveness)) OR (tw:(capacidade de resposta,)) OR (tw:(sensibilidad))))

**#1 AND #2 AND #3 - CINAHL e Dentistry & Oral Sciences Source (via EBSCO)
(28.08.2020)**

#1

(((TX "Dental caries" OR AB "Dental caries" OR SU "Dental caries" OR TX Carie\$ OR AB Carie\$ OR TX "Streptococcus mutans" OR AB "Streptococcus mutans" OR SU "Streptococcus mutans" OR TX Cariogenic\$ OR AB Cariogenic\$ OR TX Cariogenicity OR AB Cariogenicity) OR TX "Dental Decay" OR AB "Dental Decay" OR TX "Tooth Decay" OR AB "Tooth Decay" OR TX Decay OR AB Decay OR TX Carious OR AB Carious OR TX "Dental caries risk" OR AB "Dental caries risk") OR TX Dentistry OR AB Dentistry OR SU Dentistry)

#2

((((TX "Nutrition Assessment" OR AB "Nutrition Assessment" OR SU "Nutrition Assessment" OR TX "Dietary questionnaire" OR AB "Dietary questionnaire" OR TX "Food Frequency Questionnaire" OR AB "Food Frequency Questionnaire" OR TX "food assessment" OR AB "food assessment" OR TX "Health Impact Assessment" OR AB "Health Impact Assessment" OR SU "Health Impact Assessment") OR TX "24h Recalls" OR AB "24h Recalls" OR TX "Diet evaluation" OR AB "Diet evaluation" OR TX "Feeding Behavior" OR AB "Feeding Behavior" OR SU "Feeding Behavior" OR TX "Consumption analysis" OR AB "Consumption analysis") OR TX Diet\$ OR AB Diet\$ OR TX "Food questionnaire" OR AB "Food questionnaire" OR TX "Food frequency" OR AB "Food frequency" OR TX "Food consumption" OR AB "Food consumption" OR TX "food diary" OR AB "food diary") OR TX "Sugar ingestion" OR AB "Sugar ingestion" OR TX "Sugar ingestion" OR AB "Sugar ingestion") OR TX Free sugar\$ OR AB Free sugar\$)

#3

((((TX Valid\$ OR AB Valid\$ OR TX "Validation Study" OR AB "Validation Study" OR TX "Validity of result\$" OR AB "Validity of result\$" OR TX "Reproducibility of Results" OR AB "Reproducibility of Results" OR SU "Reproducibility of Results" OR TX Validity OR AB Validity) OR TX "Validation process" OR AB "Validation process" OR TX Validating OR AB Validating OR TX "Reproducibility" OR AB "Reproducibility" OR TX "Reliable" OR AB "Reliable" OR TX ("Sensitivity and Specificity") OR AB ("Sensitivity

and Specificity") OR SU ("Sensitivity and Specificity")) OR TX Sensitivity OR AB Sensitivity OR TX Specificity OR AB Specificity OR TX Unreliable OR AB Unreliable OR TX Questionnaire\$ OR SU Questionnaire\$ OR TX Feasibility OR AB Feasibility) OR TX Feasibility OR AB Feasibility OR TX Instrument\$ OR AB Instrument\$ OR TX Measure OR AB Measure OR TX Interpretability OR AB Interpretability OR TX Responsiveness OR AB Responsiveness)

Eligibility criteria

Studies on the construction and/or validation of instruments/scales to assess food consumption for use in dentistry, with or without evaluation of their psychometric properties, were included. Review studies, letters to the editors, and research protocols were excluded.

Selection of studies

The included articles were initially selected by reading the titles and abstracts, and duplicate studies from both the databases were eliminated. A complete reading was performed if there was any doubt regarding the eligibility of the study due to lack of information in the abstract. All included studies were read in full to confirm the eligibility criteria. The entire process was performed independently by two examiners (C. S and S. V. R), who strictly complied with the inclusion and exclusion criteria. Disagreements were resolved through discussion with the help of a third reviewer (J.S.R). The agreement among reviewers was evaluated using the Kappa coefficient; it was acceptable when the values were above 0.7 to 0.8 (substantial agreement) and 0.8 to 1.0 (excellent agreement) (Landis and Koch 1977). The value found after reading the titles and abstracts was 0.77, which was considered a substantial agreement. This value increased to 0.81 after reading the full text, which was considered an excellent agreement [Landis and Koch 1977].

Data extraction

Two reviewers independently performed data extraction (C.S and S.V.R) using an extraction form designed for this review, according to the characteristics identified in the studies. Before extraction, the form was pre-tested, and the researchers were trained. Disagreements among researchers were resolved through discussions with the help of a third reviewer (J.S.R).

Quality assessment

The methodological quality of the included studies was assessed using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) [Mokkink, de Vet et al. 2018], which is a specific instrument for validation studies. This instrument allowed the evaluation of the quality of studies by considering the following psychometric properties: internal consistency, test-retest reliability, content validity, face validity, and construct validity [Campos, Felicidade et al. 2019]. It had 10 assessment categories that were used to determine whether the study met the standards of good quality, response theory, and generalizability of results [Prudêncio, Serafim et al. 2021].

Categories 1 and 2 referred to the validity of the content (requirements, general scope of the project, and validation by specialists); categories 3–5 referred to structural validity, internal consistency (CI), and cross-cultural validity, respectively; together, they formed the internal structure. Categories 6–10 formed the properties of the remaining measurements: reliability, measurement error, hypothesis testing for construct validity, and responsiveness (Melo, Alves et al. 2014). These categories consisted of items that assessed the measurement properties. Each item can be classified as (1) very good, (2) suitable, (3) doubtful, (4) inappropriate, or (5) not applicable (NA). The final score for each category was determined using the lowest ranked option [Prudêncio, Serafim et al. 2021].

The risk of bias assessment was performed independently by two examiners (C.S and S.V.R). Differences were discussed until a consensus was reached, with the participation of a third reviewer (J.S.R.). All researchers were trained to use COSMIN. Inter-rater agreement was calculated using the intraclass correlation coefficient (ICC) (Campos, Felicidade et al. 2019). For each property, the percentage of items classified as "excellent", "very good", "adequate", "doubtful, or "inadequate" was calculated. For this, Wilcoxon's non-parametric test for ordinal variables was performed to assess whether there was any systematic error; the significance level adopted was $p \geq 0.05$, with no significant difference between the mean scores of the classification of each examiner. The ICC was 0.92, indicating an excellent inter-rater agreement. Wilcoxon's non-parametric test resulted in a value of $p = 0.166$; therefore, there was no systematic error among the evaluators.

Data analysis and synthesis

Data analysis was performed considering the following items: author, country, instrument language, sample, assessment instrument, instrument structure, type of food, comparison with clinical condition, adaptation of an existing instrument, purpose of the study, and psychometric properties.

It was not possible to conduct a meta-analysis of the psychometric properties as only the studies by Psoter et al. [Psoter, Gebrian et al. 2008] and Devenish et al. [Devenish, Mukhtar et al. 2017] expressed the Pearson correlation data. The structures of these instruments were very discrepant; one was a FFQ with evaluation of foods and beverages with total and free protective sugars [Devenish, Mukhtar et al. 2017] and the other was a FFQ regarding the consumption of sugary drinks and candies, and the assessment of the amount of added sugars in food preparation [Psoter, Gebrian et al. 2008].

RESULTS

The article selection process is illustrated in Figure 1. After searching the databases, gray literature, and removing duplicate studies, a total of 6382 articles were found. After reading the titles and abstracts, 8 articles were selected; however, at the stage of reading the texts in full, one study was excluded because it was not a specific instrument for assessing food consumption and the risk of caries. Finally, 7 articles were included in this literature review.

Table 2 summarizes the information found in the included articles [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Devenish, Mukhtar et al. 2020; Patenaude, Papagerakis et al. 2020]. Most of the studies were conducted in the American continent, United States [Custodio-Lumsden, Wolf et al. 2016], Canada [Patenaude, Papagerakis et al. 2020] and Haiti [Psoter, Gebrian et al. 2008]. Two studies were carried out in Asia, Japan [Shinga-Ishihara, Nakai et al. 2014] and Malaysia [Ngah, Hasan et al. 2019] and two in Oceania, both in Australia [Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019]. The instruments available in the literature are in English [Custodio-Lumsden, Wolf et al. 2016; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Patenaude, Papagerakis et al. 2020], Japanese [Shinga-Ishihara, Nakai et al. 2014], Creole [Psoter, Gebrian et al. 2008], and Malay [Ngah, Hasan et al. 2019]. Only one study performed translation and cross-cultural validation of an instrument [Shinga-Ishihara, Nakai et al. 2014], whereas the others were construction studies. Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014] built and adapted the instrument from an English version [Evens 1997] to Japanese. The English version of the study [Evens 1997] was not found for complete reading, preventing the quality and validation of the original instrument from being evaluated, considering its psychometric properties.

Complete validation of the instrument, considering the categories evaluated by COSMIN, was performed only by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014]. Five studies were identified [Psoter, Gebrian et al. 2008; Custodio-Lumsden, Wolf et al. 2016; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020] that analyzed at least one psychometric

property (linguistic validation, test-retest reliability and reproducibility, expert panel, pilot test, criterion validity, and internal consistency). Only one study did not assess any psychometric property [Devenish, Mukhtar et al. 2017].

Regarding the target population of the instruments, those proposed by Amezdroz et al. [Amezdroz, Carpenter et al. 2019] and Custodio-Lumsden et al. [Custodio-Lumsden, Wolf et al. 2016] were intended for children but answered by parents, as they evaluated the diet of children up to preschool age [Custodio-Lumsden, Wolf et al. 2016; Amezdroz, Carpenter et al. 2019]. The instrument built by Psoter et al. [Psoter, Gebrian et al. 2008] performed a joint dietary assessment of parents and their children. Four studies validated their instruments in an adult population and Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014] applied the Japanese version of the FFQ to pregnant women [Shinga-Ishihara, Nakai et al. 2014]. The studies by Ngah et al. [Ngah, Hasan et al. 2019] and Devenish et al. [Devenish, Mukhtar et al. 2017] obtained information on the frequency of food consumption from parents of children to assess the family's habitual consumption. Only Patenaude et al. [Patenaude, Papagerakis et al. 2020] validated their instrument in the general population.

Regarding the objectives of the instruments, some evaluated the frequency of general food consumption while others focused on foods with high sugar content or associated with other measures. Those with more than one measure were evaluated in addition to food frequency, family behavior, or other risk factors related to the development of dental caries. Devenish et al. [Devenish, Mukhtar et al. 2017] evaluated general food consumption, considering not only cariogenic foods but also foods that protected general health. Three studies [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Devenish, Mukhtar et al. 2017] evaluated food frequency related only to sugary food. Two studies developed an FFQ using other methods [Custodio-Lumsden, Wolf et al. 2016; Patenaude, Papagerakis et al. 2020]. The instrument proposed by Patenaude et al. [Patenaude, Papagerakis et al. 2020] evaluated the frequency of food consumption with eating behavior. Custodio-Lumsden et al. [Custodio-Lumsden, Wolf et al. 2016] developed a tool called MySmileBuddy, which evaluated dietary practices, caregiver attitudes and beliefs, fluoride use, and family history, in addition to food consumption. Finally, the

instrument by Amezdroz et al. [Amezdroz, Carpenter et al. 2019] was a cariogenic scale that measured the cariogenicity of foods and beverages. Only the studies by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014], Custodio-Lumsden et al. [Custodio-Lumsden, Wolf et al. 2016] and Amezdroz et al. [Amezdroz, Carpenter et al. 2019] compared the results of the proposed instrument with the patient's clinical condition. These studies compared the frequency of food consumption with the following clinical parameters: presence of caries using clinical oral examinations (ICDAS II) [Amezdroz, Carpenter et al. 2019] and assessment of the level of salivary mutans streptococci [Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016].

Table 3 presents the classification of the quality of the studies for each psychometric property considered in the COSMIN checklist [Mokkink, de Vet et al. 2018]. Regarding the general evaluation of the studies, all were classified as “inadequate” [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020]. Overall, among all the categories evaluated by COSMIN, the category “3” of internal structure and structural validity (3) was the most frequently evaluated as adequate. The construct validity category (9) was the worst evaluated, as it was not considered by most studies and only were classified as “adequate” [Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016]. The study by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014] received the best evaluation, since it was the only study to evaluate all the psychometric properties used in COSMIN; however, it also had a general classification of “inadequate” as not all criteria for each category were met.

All studies evaluated the reliability categories (test–retest) [Psoter, Gebrian et al. 2008; Custodio-Lumsden, Wolf et al. 2016; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020]; however, only one study (Shinga-Ishihara, Nakai et al. 2014) obtained an “adequate” quality rating in this category. With respect to responsiveness, the studies by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014], Devenish. et al. [Devenish, Mukhtar et al. 2017], and Custodio-Lumsden et al. [Custodio-Lumsden, Wolf et al. 2016] were

classified as “doubtful”. The others presented “inadequate” quality for this category [Psoter, Gebrian et al. 2008; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020], as no study had fully validated this psychometric property.

Figure 1. Flowchart of the study selection process according to the PRISMA 2020 guidelines.

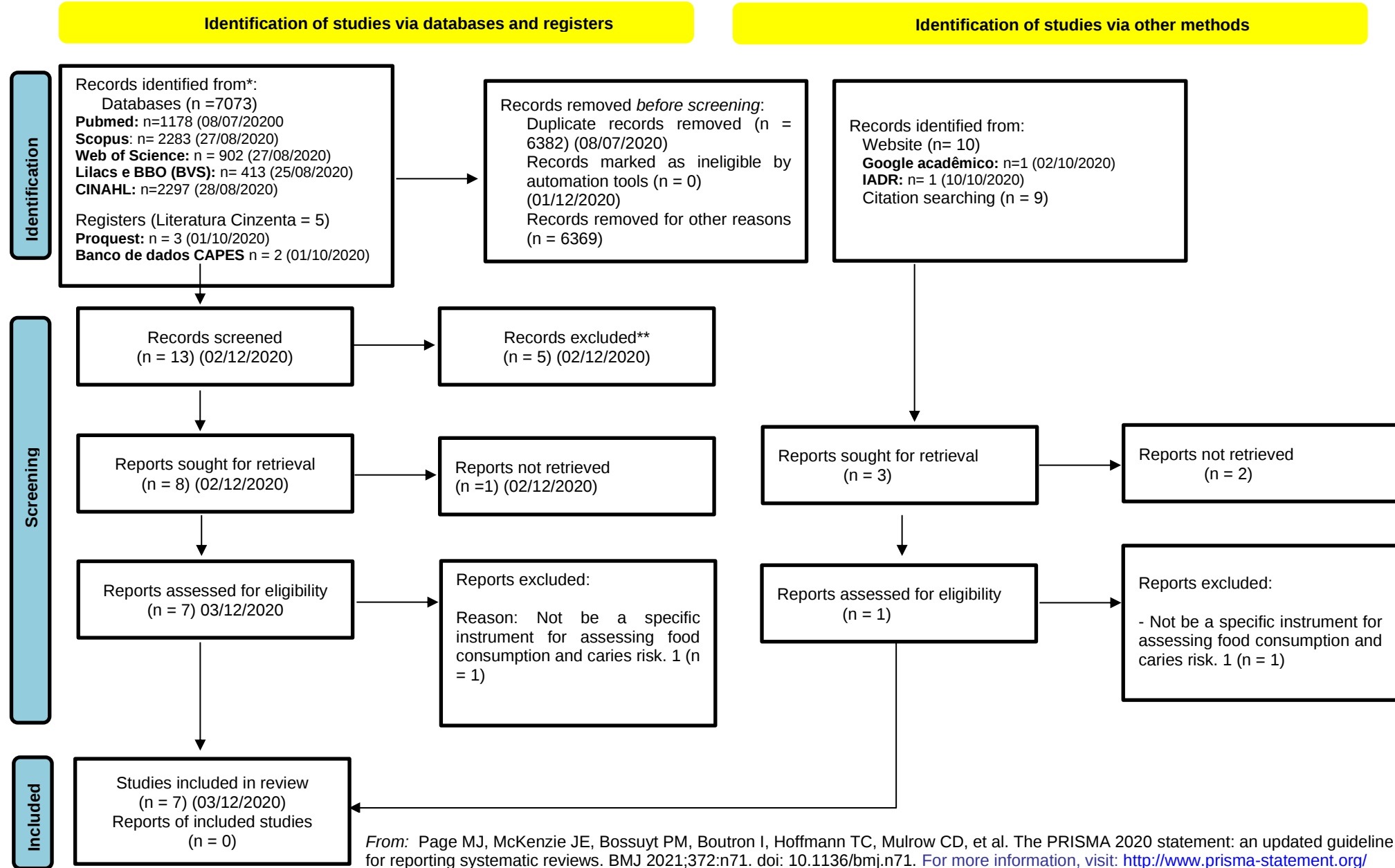


Table 2. Summary of studies included in this systematic review.

Autor	País	Idioma do instrumento	Amostra	Instrumento de avaliação	Estrutura do instrumento	Tipo de alimentos	Comparação com condição clínica	Adaptação de instrumento existente	Finalidade do estudo	Propriedades psicométricas avaliadas
Amezdroz, et al,2019 ¹	Austrália	English	Child (parent interview) 6 months n= 377 12 months n= 330 18 months n= 387 5 years n = 259	Children's usual consumption: FFQ/Karyogenic Scale	32 food items / 8 answers/ Nutritional Composition	Food and drink according to the degree of cariogenicity	Clinical Oral Examinations (ICDAS II)	No	Construction and Validation	Construct validity; pre-test; Reproducibility (test retest)
Patenaude, et al,2020 ⁶	Canadá	English	General population Female n= 67 Male n= 23	FFQ/ Eating Behavior	7 food frequency blocks / 23 food items	Foods with fermentable carbohydrates and with direct sugar action	No	No	Construction and Validation	Specialists; pilot test; Reliability (test retest)
Psoter, et al, 2008 ¹⁷	Haiti	Creole	Mother and son (9-17 years) Mothers n=30 Children n=30	FFQ of sugar consumption and added sugar in food preparation	5 questions/child (sugar consumption) 5questions/mother (amount of added sugar)	Drinks and candys	No	Yes	Validation	Reliability (test retest), pre-test, content validity, translation
Ngah, et al,2019 ²⁰	Malásia	Malay	Parents of children (6 to 11 years old) Female n= 40 Male n=9	FFQ Cariogenic and oral health	English Version translated to Bahasa Melayu 15 food and beverage	Foods and drinks cariogenic industrialized	No	Yes	Language Validation	Validation Linguistics; reliability (test retest); pilot test, experts; contents

Autor	País	Idioma do instrumento	Amostra	Instrumento de avaliação	Estrutura do instrumento	Tipo de alimentos	Comparação com condição clínica	Adaptação de instrumento existente	Finalidade do estudo	Propriedades psicométricas avaliadas
					categories/26 food items					
Shinga-Ishihara, et al,2014 ²¹	Japan	Japanese	pregnant women (19 to 43 years old) n=355	FFQ Sugary and industrialized	FFQ translated Japanese version - 38 food items.	Foods and drinks related to caries	Salivary streptococci mutans Dentocult SM Strip Mutans level (Orion Diagnostica, Espoo, Finland)	Yes	Cross-Cultural Validation	validation of construct; Internal consistency; Reliability (test-retest) and Intra-examiner, criterion; translational; pilot test; contents
Custodio-Lumsden et al, 2015 ²⁴	USA	English	Children (2 to 6 years) Female n=96 Male n=12	MySmileBuddy (MSB)	NA	NA	Salivary mutans streptococci level (saliva sample)	No	Validation	Reliability (test-retest) Intra-examiner; criterion and construct validity
Devenish, et al ¹⁸	Australia	English	Parents of children (18 to 36 months) n=97	FFQ	FFQ 24 blocks/89 food items/4 descriptive questions	Foods and beverages with total, sugar-free, and protective content	No	No	Construction	NA

NA= Not applicable

Table 3. Study quality rating for each of the psychometric properties according to the COSMIN Checklist.

Autor/Ano	Categoria 1	Categoria 2	Categoria 3	Categoria 4	Categoria 5	Categoria 6	Categoria 7	Categoria 8	Categoria 9	Categoria 10	Final
Amezdroz et al. 2019 ¹	Inappropriate	NA	NA	NA	NA	Inappropriate	Inappropriate	Inappropriate	Inappropriate	NA	Inappropriate
Patenaude <i>et al.</i> 2020 ⁶	Inappropriate	Inappropriate	Doubtful	NA	NA	Inappropriate	Inappropriate	Inappropriate	NA	NA	Inappropriate
Psoter <i>et al.</i> 2008 ¹⁷	Inappropriate	Inappropriate	Adequate	Inappropriate	Doubtful	Doubtful	Inappropriate	Inappropriate	NA	NA	Inappropriate
Ngah et al. 2019 ²⁰	Inappropriate	Doubtful	Adequate	Inappropriate	Adequate	Inappropriate	Doubtful	Inappropriate	NA	NA	Inappropriate
Shinga-Ishihara et al. 2014 ²¹	Inappropriate	Doubtful	Adequate	Adequate	Adequate	Adequate	Adequate	Inappropriate	Adequate	Doubtful	Inappropriate
Custodio-Lumsden 2016 ²⁴	NA	Inappropriate	Inappropriate	Inappropriate	NA	Inappropriate	Inappropriate	Inappropriate	Adequate	Doubtful	Inappropriate
Devenish, et al. 2019 ¹⁸	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Inappropriate

NA= Not applicable. Category 1: General requirements (Cognitive interview/pilot test) Category 2: Content Validity (Experts) Category 3: Internal structure: Structural validity Category 4: Internal structure: Internal consistency Category 5: Internal structure: Cross-cultural validity Category 6: Reliability (Test-retest) Category 7: Measurement error Category 8: Hypothesis test Category 9: Construct validity Category 10: Responsiveness Final: Final ranking of each category, determined by the lowest option.

DISCUSSION

The assessment of food consumption is a crucial part of the approach to dental treatment [Fernandes 2015], as intake of sugary foods and drinks is directly related to the development of dental caries [Tini and Long 2015; Mokkink, de Vet et al. 2018; Campos, Felicidade et al. 2019]. The correct evaluation of this consumption represents a great challenge for these professionals; therefore, the use of reliable instruments with validated psychometric properties [Campos, Felicidade et al. 2019] and the ability to effectively identify foods with potential risks for oral diseases [Fernandes 2015; Rosa 2015] is of paramount importance.

Most of the instruments found in this review are FFQs related to the consumption of foods and beverages with a high content of total and free sugars. They are aimed at the age group of children up to preschool age [Psoter, Gebrian et al. 2008; Custodio-Lumsden, Wolf et al. 2016; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019] and are in English [Psoter, Gebrian et al. 2008; Custodio • Lumsden, Wolf et al. 2016; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Patenaude, Papagerakis et al. 2020]. Regarding the quality of these instruments, all studies were classified as “inadequate” in the general assessment, with reliability (test-retest) being the most validated psychometric property [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020]; only one study [Shinga-Ishihara, Nakai et al. 2014] validated all the psychometric properties used in COSMIN, even if incompletely.

The assessment of food consumption through FFQs allows epidemiological and clinical research to be conducted to assess the role of diet in the etiology of various diseases, such as dental caries [Fernandes 2015; Rosa 2015]. The studies included the development and validation of FFQs regarding the consumption of total sugars in foods and beverages [Psoter, Gebrian et al. 2008; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020].

These instruments allowed the investigation and direct assessment of the regularity with which the patient had access to food with high cariogenic potential and the association of such consumption with clinical outcomes. They can also be used for individual patient assessment in dental clinics, as they allow

personalization of dietary advice and encourage discussion between professionals and patients [Fernandes 2015; Rosa 2015; Arheiam, Albadri et al. 2016; Arheiam, Albadri et al. 2018]. An important finding of this study was the predominance of instruments in English, with only three in other languages: Japanese, Creole, and Malay [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Ngah, Hasan et al. 2019].

However, diet assessment has been carried out in studies in the field of dentistry, in different countries and languages [Arheiam, Brown et al. 2016; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Hancock, Zinn et al. 2020; Mahboobi, Pakdaman et al. 2021], as it is one of the factors related to dental caries. Regarding the diet assessment methods most frequently used in published studies, two systematic reviews [Moynihan and Kelly 2014; Hancock, Zinn et al. 2020; Mahboobi, Pakdaman et al. 2021] found that most studies used the DA method, followed by the FFQ, to assess the relationship between dental caries and food consumption. The main finding of these prospective studies, with populations of children and adolescents, was that high-frequency consumption of sugary foods was associated with a greater occurrence of dental caries [Hancock, Zinn et al. 2020; Mahboobi, Pakdaman et al. 2021].

The FFQ, DA, and 24-hour recall (R24h) are the reference methods to be used in epidemiological studies and/or in clinical practice to assess food consumption [Fernandes, Moura et al. 2013; Fernandes 2015; Rosa 2015]. To choose the best method, objective of the analysis (individual or epidemiological), psychometric properties, and possibility of adapting the measure for the study population should be considered, considering the regional diversity of eating habits [Evens 1997; Shinga-Ishihara, Nakai et al. 2014]. Both DA and/or 24hR can be applied to any population; however, for epidemiological studies, these instruments do not encompass variability, habits, and seasonality of the diet, and do not represent habitual consumption [Fisberg, Marchioni et al. 2009]. In addition, the compilation and analysis of their results are more complex because of the qualitative nature of the data, which is difficult to apply in large epidemiological studies [Fernandes 2015; Pedraza and Menezes 2015]. In contrast, the FFQ estimates dietary consumption over time based on the possibility of measuring the intensity of exposure and classifying individuals according to the frequency of consumption. This makes it possible to estimate

the association of consumption categories with the development of development of chronic non-communicable diseases [Fernandes 2015; Rosa 2015].

However, as these instruments assess long-term food consumption, memory bias and underestimation of the absolute amount of consumption may occur [Fernandes 2015]. This variability of methods found in the literature can make comparisons between studies difficult as some of them, such as DA and R24h, do not allow long-term evaluation. Considering the complexity of obtaining valid measurements of food intake in epidemiological studies, there is a need to adapt and validate instruments [Kolodziejczyk, Merchant et al. 2012; Fernandes 2015]. In addition, the formulation of instruments with the help of nutrition professionals plays an important role in the reliable construction and evaluation of food consumption.

Most of the available FFQs were constructed and validated to assess the frequency of consumption of sugary foods and beverages and their degree of cariogenicity in children up to preschool age [Psoter, Gebrian et al. 2008; Custodio • Lumsden, Wolf et al. 2016; Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019]. Assessing food consumption at this age is relevant in dentistry as dental caries can evolve rapidly, and any sign of injury before the age of three is already considered severe caries. The high consumption of sugar in the first years of life can be explained by several factors occurring simultaneously [Yee, Lwin et al. 2017]. Parents' expectations regarding their children's food quantity and frequency may be at odds with their children's physiological needs; as there is a greater resistance to food consumption at this age, parents incentivize increased food consumption by offering sweets. It should also be considered that eating habits reflect cultural and symbolic patterns; offering of cariogenic products can have the meaning of affection and love [Fernandes 2015; Rosa 2015; Yee, Lwin et al. 2017].

Even in the age group of children up to preschool age, food choices and frequency are totally dependent on parents and caregivers [Scaglioni, De Cosmi et al.], and children do not have the cognitive ability to fully self-report their food consumption; therefore, parents and caregivers must answer these questions [Devenish, Mukhtar et al. 2017; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019]. There are limitations to obtaining information about a child's nutrition from parents and caregivers, since most of them cannot stay with their children

for most of the day and information can be hidden [Kolodziejczyk, Merchant et al. 2012; Morikava, Fraiz et al. 2018]. However, these instruments are still relevant given the importance of identifying risk factors for caries, including diet [Fernandes 2015; Rosa 2015]. This reinforces the importance of evaluating the construct validity of these instruments, especially convergent validity, by comparing their results with patients' clinical conditions. In the literature, there are few instruments validated for adults and the elderly [Moynihan and Kelly 2014]. Once the disease burden reaches adulthood, it is important that the relationship between diet and dental caries be investigated in these groups [Moynihan and Kelly 2014; Hancock, Zinn et al. 2020; Mahboobi, Pakdaman et al. 2021].

The study by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014] was the only one that translated and cross-culturally adapted an existing instrument in English into Japanese, designed to be applied to children [Evens 1997]. The original version had 35 food items with assessment of food consumption related to caries during the previous month [Evens 1997]; however, this instrument was not found in its complete version, so its psychometric properties could not be evaluated. The Japanese version [Shinga-Ishihara, Nakai et al. 2014] added foods that were regularly consumed in the country to increase the validity of the content. In this version, the questionnaire was validated exclusively for the population of pregnant women as dietary preferences often change during this period and can lead to frequent ingestion of cariogenic foods [Shinga-Ishihara, Nakai et al. 2014]. Studies by Psoter et al. [Psoter, Gebrian et al. 2008] and Ngah et al. [Ngah, Hasan et al. 2019] only translated the instrument, being translated/back-translated from a source language to another language using native speakers; it was not reported whether cross-cultural adaptation was carried out in the test-retest version. The study by Psoter et al. [Psoter, Gebrian et al. 2008] only validated reproducibility (test-retest). It was not found in its original version, As the references cited in this instrument were not accessible, it cannot be confirmed if construction and psychometric properties were evaluated due to lack of access to the references cited in the instrument.

For cross-cultural validation, adaptations must be made carefully, since the instrument must be internally reliable; for international comparisons, ensuring the quality of the data and the information produced is important. The instrument must be translated/back-translated from a source language to another language

using native speakers. Mastery of the other language, carried out by peers in a hidden way, carrying out the comparability of the translation results, adaptation of the culture, habits, and seasonality of the diet, and carrying out a pre-test for necessary modifications [Shinga-Ishihara, Nakai et al. 2014], demonstrates reliability and validity [Shinga-Ishihara, Nakai et al. 2014; Ngah, Hasan et al. 2019]. In the study by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014], all stages of translation and cross-cultural adaptation were carried out, ensuring reliability and comparability with the original study [Ngah, Hasan et al. 2019]; however, there were flaws in the validation process as the criteria for each category were not fully complied with and therefore, the overall rating of the study was “inadequate”. Three studies [Shinga-Ishihara, Nakai et al. 2014; Custodio • Lumsden, Wolf et al. 2016; Amezdroz, Carpenter et al. 2019] performed construct validation and compared the results of the proposed instrument with the clinical condition of the patients. In the studies by Custodio-Lumsden et al. [Custodio-Lumsden, Wolf et al. 2016] and Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014] salivary mutans streptococcal levels showed a statistically significant relationship with dietary intake, indicating that their higher levels were associated with higher scores for cariogenic food intake. The study by Amezdroz et al. [Amezdroz, Carpenter et al. 2019], with the use of clinical oral examinations (ICDAS II), showed a significant increase in carious lesions as advanced disease at the age of five years, when the child’s intake of discretionary foods that do not provide nutrients, including a cariogenic diet, increases. These relationships between the instrument and the clinical condition are essential for findings to be effectively generalized, and for the information to reflect the individual’s real condition and be able to measure exactly what is proposed [Fernandes 2015].

The most frequently evaluated property in the studies was reliability (test-retest) [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020]. Repetitions of measurements were used to assess the reliability of the instrument, which showed the level of correlation between the distributions of scores in the same test on two different and independent occasions [Alexandre, Rodrigues et al. ; Cunha, Pereira de Almeida Neto et al. 2016; Alves 2017]. If the correlation between the results of

the two applications was significantly positive, the instrument could be considered reliable [Cunha, Pereira de Almeida Neto et al. 2016]. Despite being highly evaluated, only one study [Shinga-Ishihara, Nakai et al. 2014] obtained an “adequate” quality rating according to COSMIN. The studies [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016; Amezdroz, Carpenter et al. 2019; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020] validated this property incompletely, with the absence of an interval between two applications, lack of description of the training, and whether the interviewers were able to carry out the research, making the quality of the measurement instrument questionable [Cunha, Pereira de Almeida Neto et al. 2016; Alves 2017]. In addition, despite being a necessary quality, it must be evaluated together with other components within the psychometric properties to guarantee its accuracy [Cunha, Pereira de Almeida Neto et al. 2016; Campos, Felicidade et al. 2019]. The COSMIN checklist is a tool that aims to improve the methodological quality of studies on property measures [Mokkink, de Vet et al. 2018]. The general results found showed that all studies included in this review were classified as “inadequate,” since the general classification of the quality of each study was always determined by the worst condition found in the criteria evaluated in each category. The study by Shinga-Ishihara et al. [Shinga-Ishihara, Nakai et al. 2014] received the best assessment according to the COSMIN scale, as it was the only one to assess all predicted psychometric properties. However, it was still classified as inadequate because the hypothesis testing and response capacity categories were incomplete [Shinga-Ishihara, Nakai et al. 2014]. This showed that most studies have important limitations that influence the evidence on the instruments if they measure what they purport to measure [Vermelho 2014]. The rigidity of the COSMIN general classification was to ensure an instrument’s validity, reliability, and reproducibility; the psychometrics of each measure must be carefully considered [Cunha, Pereira de Almeida Neto et al. 2016; Campos, Felicidade et al. 2019; Prudêncio, Serafim et al. 2021].

It is possible to highlight other limitations that affected the methodological quality of the study, such as the limitation of the sample population and failure to detail all psychometric properties performed [Psoter, Gebrian et al. 2008; Shinga-Ishihara, Nakai et al. 2014; Custodio-Lumsden, Wolf et al. 2016; Ngah, Hasan et al. 2019; Patenaude, Papagerakis et al. 2020]. Among the main flaws, one study

[Devenish, Mukhtar et al. 2017] did not have any psychometric properties for validation. When a psychometric instrument is not tested, its measurement is imprecise, which limits the validity of the inferences obtained from this instrument [Alves 2017]. It is necessary to compare these instruments with the clinical conditions of oral health to ensure that they are measuring exactly what is expected [Fernandes 2015; Cunha, Pereira de Almeida Neto et al. 2016].

Regarding the limitations of this systematic review, it is important to note that it was not possible to have full access to two studies [Evens 1997; Shinga-Ishihara, Nakai et al. 2014], which may have limited the coverage of this review. However, all possible attempts were made to access complete articles.

CONCLUSION

Most of the instruments found in this literature review consisted of FFQs related to the consumption of foods and beverages with a high total and free sugar content, aimed at the age group of children of preschool age. All studies were classified as “inadequate” in the general assessment. Reliability (test-retest) was the most validated psychometric property, and only one study validated all psychometric properties used by COSMIN. Advances are needed in the validation process of these instruments, with appropriate and complete methods for validating their psychometric properties, to minimize measurement errors and avoid incorrect estimations.

Statements

Statement of ethics

There are no conflicts of interest to declare. There was no need to go through the Research Ethics Committee to deal with a systematic literature review.

Funding Sources

This study was funded in part by the Integrated Research Center (CIP) of the Pontifical Catholic University of Paraná, Brazil.

Author Contributions

All authors have reviewed and approved the complete manuscript.

Caroline Souza dos Santos, Juliana Schaia Rocha, Samuel Jorge Moysés conceived the ideas of the systematic review and the methodological approach; Caroline Souza dos Santos, Juliana Schaia Rocha performed the screening and data extraction of this literature review, the final selection was performed by Caroline Souza dos Santos, Juliana Schaia Rocha Rodrigo Nunes Rached, Marcia Helena Baldani Pinto, Renata Iani Werneck. Gil Guilherme Gasparello helped with the English language and the led the final writing of the manuscript.

All the author declare that this manuscript is entirely original and has not been reviewed for publication elsewhere. There was no overlap with the other manuscripts under review. The authors assume responsibility for the content of the manuscript, including reviewing and approving this version and complying with authorship requirements.

REFERENCES

Amezdroz E, Carpenter L, Johnson S, Flood V, Dashper SG, Calache H, et al. Feasibility and development of a cariogenic diet scale for epidemiological research. *Int J Paediatr Dent*. 2019;29(3):310-24.

Alexandre N, Rodrigues R, Lima M, Gallasch C. Propriedades de medida dos instrumentos de avaliação do equilíbrio e risco de quedas na doença de Parkinson: uma revisão sistemática - UFMG. 2014.

Alves IB. Validade e confiabilidade do questionário nórdico de sintomas musculoesqueléticos: uma revisão sistemática de literatura: Universidade Federal da Bahia Faculdade de Medicina da Bahia. UFBA. 2017.

American Dental Association 2022. *Department of Scientific Information, Evidence Synthesis & Translation Research, ADA Science & Research Institute, LLC*. [Citação em 11 de outubro de 2011.] Disponível a partir de: <https://www.ada.org/resources/research/science-and-research-institute/oral-health-topics/nutrition-and-oral-health>.

Arheiam A, Albadri S, Brown S, Burnside G, Higham S, Harris R. Are diet diaries of value in recording dietary intake of sugars? A retrospective analysis of completion rates and information quality. *Br Dent J*. 2016;221(9):571-6.

Arheiam A, Albadri S, Lavery L, Harris R. Reasons for low adherence to diet-diaries issued to pediatric dental patients: a collective case study. *Patient Preference Adherence*. 2018;12:1401-11.

Arheiam A, Brown SL, Burnside G, Higham SM, Albadri S, Harris RV. The use of diet diaries in general dental practice in England. *Community Dent Health*. 2016;33(4):267-73.

Brasil, Ministério da Saúde, Secretária de Atenção a Saúde, Departamento de Atenção Básica - Saúde Bucal/ Ministério da Saúde, Secretária de Atenção a Saúde, Departamento de Atenção Básica - Brasília: Ministério da Saúde, 2008. 92p –

Campos M, Felicidade C, Aguiar S, Vieira D. Propriedades psicométricas de questionários de atividade física na adolescência: revisão sistemática. *Revista Brasileira de Atividade Física & Saúde*. 2019;23:1-13.

Cunha C, Pereira de Almeida Neto O, Stackfleth R. Principais métodos de avaliação psicométrica da confiabilidade de instrumentos de medida. *Revista Brasileira Ciências da Saúde - USCS*. 2016;14:75-83.

Custodio Lumsden CL, Wolf RL, Contento IR, Basch CE, Zybert PA, Koch PA, et al. Validation of an early childhood caries risk assessment tool in a low-income Hispanic population. *J Public Health Dent*. 2016;76(2):136-42.

Devenish G, Mukhtar A, Begley A, Do L, Scott J. Development and Relative Validity of a Food Frequency Questionnaire to Assess Intakes of Total and Free Sugars in Australian Toddlers. *Int J Environ Res Public Health*. 2017;14(11)1361.

Devenish G, Mukhtar A, Begley A, Spencer AJ, Thomson WM, Ha D, et al. Early childhood feeding practices and dental caries among Australian preschoolers. *Am J Clin Nutr*. 2020;111(4):821-8.

Evens CC. Snacking patterns as a risk factor for early childhood caries. Proquest: Universidade de Washington, ProQuest Dissertations Publishing, 1997.

Fernandes MLdMF, Moura FMP, Gamaliel KS, Corrêa-Faria P. Cárie dentária e necessidade de tratamento ortodôntico: impacto na qualidade de vida de escolares. *Pesqui bras odontopediatria clín integr*. 2013;13(1)37-43.

Fernandes MP. Construção e validação de um Questionário de Frequência Alimentar para avaliar o consumo de alimentos cariogênicos em pré-escolares 2015. 125f Dissertação (Mestrado) - Programa de Pós Graduação em Nutrição e Alimentos. UFP, Pelotas, 2015. editor. Brasil: Universidade Federal de Pelotas; 2015. 126p.

Fisberg R, Marchioni D, Colucci A. Avaliação do consumo alimentar e da ingestão de nutrientes na prática clínica. *Arquivos Brasileiros De Endocrinologia E Metabologia - Arq Bras Endocrinol Metabol*. 2009;53/5.

Hancock S, Zinn C, Schofield G. The consumption of processed sugar- and starch-containing foods, and dental caries: a systematic review. *Eur J Oral Sci*. 2020;128(6):467-75.

Kolodziejczyk JK, Merchant G, Norman GJ. Reliability and validity of child/adolescent food frequency questionnaires that assess foods and/or food groups. *J Pediatr Gastroenterol Nutr*. 2012;55(1):4-13.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.

Mahboobi Z, Pakdaman A, Yazdani R, Azadbakht L, Montazeri A. Dietary free sugar and dental caries in children: A systematic review on longitudinal studies. *Health Promot Perspect*. 2021;11(3):271-80.

Melo BM, Alves AD, Nogueira EG, Jardim DN, Milhomem DG, Pereira YC, editors. A importância da prevenção baseada no IHOS. Janeiro de 2014; 1(2):14.

Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171-9.

Moreira PR. Práticas alimentares relacionadas à cárie dentária: uma revisão. Sul UFRGS, editor. Universidade Federal do Rio Grande do Sul: UFRGS; 2016. 42 p.

Morikava FS, Fraiz FC, Gil GS, de Abreu MHNGe, Ferreira FM. Healthy and cariogenic foods consumption and dental caries: A preschool-based cross-sectional study. *Oral Diseases*. 2018;24(7):1310-7.

Moynihan PJ, Kelly SA. Effect on caries of restricting sugars intake: systematic review to inform WHO guidelines. *J Dent Res*. 2014;93(1):8-18.

Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol*. 2018;18(1):5.

Ngah R, Hasan R, Rahman NA. Linguistic validity and reliability of cariogenic food frequency and oral health knowledge, attitude and practice questionnaires for parents of 6-11 years old children in Kota Bharu, Kelantan. *Malays J Nutr*. 2019;25(1):79-97.

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.

Patenaude SA, Papagerakis P, Lieffers JR. Development of a Nutrition Questionnaire for Dental Caries Risk Factors. *Int J Environ Res Public Health*. 2020;17,1793(5).

Pedraza DF, Menezes TN. [Food Frequency Questionnaire developed and validated for the Brazilian population: a review of the literature]. *Cien Saude Colet*. 2015;20(9):2697-720.

Prudêncio DA, Serafim TT, Marinho Mateus Lopes APSR, Maffulli N, Okubo R. Questionnaires and scales for assessment of ankle function: a systematic review of instruments translated and validated for Brazilian Portuguese. *Disabil Rehabil*. 2021;43(3):309-16.

Psoter WJ, Gebrian B, Katz RV. Reliability of a sugar consumption questionnaire for rural Haiti. *P R Health Sci J*. 2008;27(1):69-74.

Rosa QFd. Development and validation of a Food Frequency Questionnaire to assess the adolescent's consumption of foods with cariogenic and erosive potential. UFP ed. Pelotas UFd, editor. Brasil: Universidade Federal de Pelotas; 2015:74 p.

Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing Children's Eating Behaviours. *MDPI Nutrient*. 2018;10(6);706.

Shinga-Ishihara C, Nakai Y, Milgrom P, Murakami K, Matsumoto-Nakano M. Cross-cultural validity of a dietary questionnaire for studies of dental caries risk in Japanese. *BMC Oral Health*. 2014;14:1.

Tini GF, Long SM, editors. Avaliação de diários alimentares de crianças atendidas na clínica infantil de uma universidade privada de São Paulo. DOI: 10.15603/2176-1000.

Vermelho NFF. Propriedades de medida dos instrumentos de avaliação do equilíbrio e risco de quedas na doença de Parkinson: uma revisão sistemática: Universidade Federal de Minas Gerais; 2014.

Yee AZ, Lwin MO, Ho SS. The influence of parental practices on child promotive and preventive food consumption behaviors: a systematic review and meta-analysis. *Int J Behav Nutr Phys Act.* 2017;14(1):47.

Figure Legends

Fig. 1. Flowchart of the study selection process according to the PRISMA 2020 guidelines

Table Legends

Table 1. Search strategy for each database. Date of searches from July 08th 2020 to October 10th 2020

Table 2. Summary of studies included in this systematic review

Table 3. Study quality rating for each of the psychometric properties according to the COSMIN Checklist

APÊNDICES

Apêndice 1. Resultados do Teste de Wilcoxon para avaliação da existência de erro sistemático nas avaliações de qualidade.

Autor/Ano	Valor de p
Patenaude et al. 2020	0,763
Psoter et al. 2008	0,657
Ngah et al. 2019	0,564
Shinga-Ishihara et al. 2014	0,257
Amezdroz et al. 2019	0,564
Devenish et al. 2017	0,632
Custodio-Lumsden et al. 2015	0,617

*Um valor de $p > 0,05$ significa que não ocorreu erro sistemático entre as avaliações dos examinadores.

Apêndice 2. Conferência dos tópicos descritos relato do resumo, conforme o recomendado pelo checklist PRISMA for Abstract

Section and Topic	Item #	Checklist item	Reported (Yes/No)
TITLE			
Title	1	Identify the report as a systematic review.	Page:6 Yes: Line 2
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Page:6 Yes: Line 4
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Page:6 Yes: Line 6
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Page:6 Yes:Line 9
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Page:6 Yes:Line 16
Synthesis of results	6	Specify the methods used to present and synthesise results.	Page:6 Yes:Line 17
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Page:6 Yes:Line 23
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Page:6 Yes:Line 23
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	Page:7 Yes:Line 1
Interpretation	10	Provide a general interpretation of the results and important implications.	Page:6 Yes:Line 3

Section and Topic	Item #	Checklist item	Reported (Yes/No)
OTHER			
Funding	11	Specify the primary source of funding for the review.	No
Registration	12	Provide the register name and registration number.	Page:7 Yes:Line 6

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

ANEXOS

COSMIN Risk of Bias checklist

Date: December, 2017

Contact

L.B. Mokkink, PhD
VU University Medical Center
Department of Epidemiology and Biostatistics
Amsterdam Public Health research institute
P.O. box 7057
1007 MB Amsterdam
The Netherlands
Website: www.cosmin.nl
E-mail: w.mokkink@vumc.nl



How to site the COSMIN Risk of Bias Checklist

Please refer to the following studies when using the COSMIN Risk of Bias Checklist:

Mokkink, L.B., De Vet, H.C.W., Prinsen, C.A.C, Patrick, D.L., Alonso, J., Bouter, L.M., et al. COSMIN Risk of Bias checklist for systematic reviews of Patient Reported Outcome Measures. Accepted for publication in Quality of Life Research.

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., Vet, H. C., et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. Submitted.

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Vet, H. C., Westerman, M. J., Patrick, D. L., et al. COSMIN methodology for evaluating the content validity of Patient-Reported Outcome Measures: a Delphi study. Submitted.

For details on how to use the COSMIN risk of Bias checklist see 'COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) – user manual' and 'COSMIN methodology for assessing the content validity of Patient-Reported Outcome Measures (PROMs) - user manual' available from our website www.cosmin.nl.

Abbreviations used:

CTT – classical test theory

DIF – differential item functioning

IRT – Item response theory

MGCFA – multi-group confirmatory factor analysis

MI – measurement invariance

NA – not applicable

PROM – patient-reported outcome measure

1PL model – 1 parameter IRT model

2PL model – 2 parameter IRT model

Instructions

Tick the boxes that need to be completed for the article

	COSMIN Risk of Bias checklist
	Box 1. PROM development
	Box 2. Content validity
	Box 3. Structural validity
	Box 4. Internal consistency
	Box 5. Cross-cultural validity\Measurement invariance
	Box 6. Reliability
	Box 7. Measurement error
	Box 8. Criterion validity
	Box 9. Hypotheses testing for construct validity
	Box 10. Responsiveness

To assess the methodological quality of each study, i.e. assessing the risk of bias of the result of a study, the corresponding COSMIN Risk of Bias box should be completed. To determine the overall quality of a study the lowest rating of any standard in the box is taken (i.e. “the worst score counts” principle). For example, if for a reliability study one item in a box is rated as ‘inadequate’, the overall methodological quality of that reliability study is rated as ‘inadequate’. The response option ‘NA’ (not applicable) is at issue for some standards. For example, when a study on structural validity is based on CTT, the standard on IRT is not applicable and this standard should not be considered in the “worst score counts”- rating for that specific study. For standards where this option is not at issue, these cells are grey and shouldn’t be used.

Box 1. PROM development**1a. PROM design***General design requirements*

		very good	adequate	doubtful	inadequate	NA
1	Is a clear description provided of the construct to be measured?	Construct clearly described			Construct not clearly described	
2	Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?	Origin of the construct clear		Origin of the construct not clear		
3	Is a clear description provided of the target population for which the PROM was developed?	Target population clearly described			Target population not clearly described	
4	Is a clear description provided of the context of use	Context of use clearly described		Context of use not clearly described		
5	Was the PROM development study performed in a sample representing the target population for which the PROM was developed?	Study performed in a sample representing the target population	Assumable that the study was performed in a sample representing the target population, but not clearly described	Doubtful whether the study was performed in a sample representing the target population	Study not performed in a sample representing the target population (SKIP items 6-12)	

<i>Concept elicitation (relevance and comprehensiveness)</i>		very good	adequate	doubtful	inadequate	NA
6	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?	Widely recognized or well justified qualitative method used, suitable for the construct and study population	Assumable that the qualitative method was appropriate and suitable for the construct and study population, but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was suitable for the construct and study population	Method used not appropriate or not suitable for the construct or study population	
7	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience		Not applicable
8	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide		Not applicable

9	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
10	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
11	Was at least part of the data coded independently?	At least 50% of the data was coded by at least two researchers independently	11-49% of the data was coded by at least two researchers independently	Doubtful if two researchers were involved in the coding or only 1-10% of the data was coded by at least two researchers independently	Only one researcher was involved in coding or no coding	Not applicable
12	Was data collection continued until saturation was reached?	Evidence provided that saturation was reached	Assumable that saturation was reached	Doubtful whether saturation was reached	Evidence suggests that saturation was not reached	Not applicable
13	For quantitative studies (surveys): was the sample size appropriate?	≥100	50-99	30-49	<30	Not applicable

1b. <u>Cognitive interview study or other pilot test</u>		very good	adequate	doubtful	inadequate	NA
14	Was a cognitive interview study or other pilot test conducted?	YES			NO (SKIP items 15-35)	
<i>General design requirements</i>						
15	Was the cognitive interview study or other pilot test performed in a sample representing the target population?	Study performed in a sample representing the target population	Assumable that the study was performed in a sample representing the target population, but not clearly described	Doubtful whether the study was performed in a sample representing the target population	Study not performed in a sample representing the target population	
<i>Comprehensibility</i>						
16	Were patients asked about the <u>comprehensibility</u> of the PROM?	YES		NO (SKIP items 17-25)	Not clear (SKIP items 17-25)	
17	Were all items tested in their final form?	All items were tested in their final form	Assumable that all items were tested in their final form, but not clearly described	Not clear if all items were tested in their final form	Items were not tested in their final form or items were not re-tested after substantial adjustments	

18	Was an appropriate qualitative method used to assess the <u>comprehensibility</u> of the PROM instructions, items, response options, and recall period?	Widely recognized or well justified qualitative method used	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was appropriate or not clear if patients were asked about the comprehensibility of the items, response options or recall period or patients not asked about the comprehensibility of the PROM instructions or the recall period	Method used not appropriate or patients not asked about the comprehensibility of the items or the response options	
19	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
20	Were skilled interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience		Not applicable
21	Were the interviews based on an appropriate interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide		Not applicable

22	Were the interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
23	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
24	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis		
25	Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM?	No problems found or problems appropriately addressed and PROM was adapted and re-tested if necessary	Assumable that there were no problems or that problems were appropriately addressed, but not clearly described	Not clear if there were problems or doubtful if problems were appropriately addressed	Problems not appropriately addressed or PROM was adapted but items were not re-tested after substantial adjustments.	Not applicable

<i>Comprehensiveness</i>		very good	adequate	doubtful	inadequate	NA
26	Were patients asked about the <u>comprehensiveness</u> of the PROM?	YES		NO or not clear (SKIP items 27-35)		
27	Was the final set of items tested?	The final set of items was tested	Assumable that the final set of items was tested, but not clearly described	Not clear if the final set of items was tested or not the final set of items was tested or the set of items was not re-tested after items were removed or added		
28	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?	Widely recognized or well justified method used	Assumable that the method was appropriate but not clearly described or only quantitative (survey) method(s) used	Doubtful whether the method was appropriate or method used not appropriate		
29	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
30	Were skilled interviewers used?	Skilled interviewers used	Interviewers had limited experience or were trained specifically for the study	Not clear if interviewers were trained or interviewers not trained and no experience		Not applicable

31	Were the interviews based on an appropriate interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide	Not applicable
32	Were the interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews or no recording and no notes	
33	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate or approach not appropriate	
34	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	

35	Were problems regarding the <u>comprehensiveness</u> of the PROM appropriately addressed by adapting the PROM?	No problems found or problems appropriately addressed and PROM was adapted and re-tested if necessary	Assumable that there were no problems or that problems were appropriately addressed, but not clearly described	Not clear if there were problems or doubtful if problems were appropriately addressed or PROM was adapted but items were not re-tested after substantial adjustments	Problems not appropriately addressed	Not applicable
----	--	---	--	--	--------------------------------------	----------------

Box 2. Content validity					
2a. Asking patients about relevance					
Design requirements		very good	adequate	doubtful	inadequate
1	Was an appropriate method used to ask patients whether each item is <u>relevant</u> for their experience with the condition?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Not clear if patients were asked whether <u>each</u> item is relevant or doubtful whether the method was appropriate	Method used not appropriate or patients not asked about the relevance of all items
2	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear	
3	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators/interviewers had limited experience or were trained specifically for the study	Not clear if group moderators/interviewers were trained or group moderators/interviewers not trained and no experience	Not applicable
4	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide	Not applicable

5	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
<i>Analyses</i>						
6	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
7	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis		

2b Asking patients about comprehensiveness						
Design requirements		very good	adequate	doubtful	inadequate	NA
8	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
9	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
10	Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience		Not applicable
11	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide		Not applicable

12	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
<i>Analyses</i>						
13	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
14	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis		

2c Asking patients about comprehensibility					
Design requirements		very good	adequate	doubtful	inadequate
15	Was an appropriate qualitative method used for assessing the <u>comprehensibility</u> of the PROM instructions, items, response options, and recall period?	Widely recognized or well justified qualitative method used	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was appropriate or not clear if patients were asked about the comprehensibility of the items, response options or recall period or patients not asked about the comprehensibility of the PROM instructions	Method used not appropriate or patients not asked about the comprehensibility of the items, response options, or recall period
16	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear	
17	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators/interviewers had limited experience or were trained specifically for the study	Not clear if group moderators/interviewers were trained or group moderators/interviewers not trained and no experience	

18	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide		Not applicable
19	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	Not applicable
<i>Analyses</i>						
20	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
21	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis		

2d. Asking professionals about relevance						
Design requirements		very good	adequate	doubtful	inadequate	NA
22	Was an appropriate method used to ask professionals whether each item is <u>relevant</u> for the construct of interest?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Not clear if professionals were asked whether <u>each</u> item is relevant or doubtful whether the method was appropriate	Method used not appropriate or professionals not asked about the relevance of all items	
23	Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included or relevant professionals were not included		
24	Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
Analyses						
25	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	

26	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	
----	---	---	--	--	--

2e. Asking professionals about comprehensiveness						
Design requirement		very good	adequate	doubtful	inadequate	NA
27	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
28	Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included or relevant professionals were not included		
29	Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
Analyses						
30	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	

31 Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	
--	---	--	--	--

Box 4. Internal consistency					
Does the scale consist of effect indicators, i.e. is it based on a reflective model? ¹ yes / no					
<i>Design requirements</i>	very good	adequate	doubtful	inadequate	NA
1 Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	Internal consistency statistic calculated for each unidimensional scale or subscale		Unclear whether scale or sub scale is unidimensional	Internal consistency statistic NOT calculated for each unidimensional scale or sub scale	
<i>Statistical methods</i>					
2 For continuous scores: Was Cronbach's alpha or omega calculated?	Cronbach's alpha, or Omega calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated	Not applicable
3 For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated	Not applicable
4 For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	SE(θ) or reliability coefficient calculated			SE(θ) or reliability coefficient NOT calculated	Not applicable
<i>Other</i>					
5 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

<i>Other</i>					
4	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws (e.g. rotation method not described)	Other important methodological flaws (e.g. inappropriate rotation method)

¹ If the scale is not based on a reflective model, unidimensionality or structural validity is not relevant.

² In a systematic review, it is helpful to make a distinction between studies where factor analysis is performed on each (sub)scale separately to evaluate whether the (sub)scales are unidimensional (unidimensionality studies) and studies where factor analysis is performed on all items of an instrument to evaluate the (expected) number of subscales in the instrument and the clustering of items within subscales (structural validity studies).

Box 3. Structural validity					
Does the scale consist of effect indicators, i.e. is it based on a reflective model? ¹ yes / no					
Does the study concern unidimensionality or structural validity? ² unidimensionality / structural validity					
<i>Statistical methods</i>	very good	adequate	doubtful	inadequate	NA
1 For CTT: Was exploratory or confirmatory factor analysis performed?	Confirmatory factor analysis performed	Exploratory factor analysis performed		No exploratory or confirmatory factor analysis performed	Not applicable
2 For IRT/Rasch: does the chosen model fit to the research question?	Chosen model fits well to the research question	Assumable that the chosen model fits well to the research question	Doubtful if the chosen model fits well to the research question	Chosen model does not fit to the research question	Not applicable
3 Was the sample size included in the analysis adequate?	FA: 7 times the number of items and ≥ 100	FA: at least 5 times the number of items and ≥ 100 ; OR at least 6 times number of items but < 100	FA: 5 times the number of items but < 100	FA: < 5 times the number of items	
	Rasch/1PL models: ≥ 200 subjects	Rasch/1PL models: 100-199 subjects	Rasch/1PL models: 50-99 subjects	Rasch/1PL models: < 50 subjects	
	2PL parametric IRT models OR Mokken scale analysis: ≥ 1000 subjects	2PL parametric IRT models OR Mokken scale analysis: 500-999 subjects	2PL parametric IRT models OR Mokken scale analysis: 250-499 subjects	2PL parametric IRT models OR Mokken scale analysis: < 250 subjects	

¹If the scale is not based on a reflective model, internal consistency is not relevant.

Box 5. Cross-cultural validity\Measurement invariance						
Design requirements		very good	adequate	doubtful	inadequate	NA
1	Were the samples similar for relevant characteristics except for the group variable?	Evidence provided that samples were similar for relevant characteristics except group variable	Stated (but no evidence provided) that samples were similar for relevant characteristics except group variable	Unclear whether samples were similar for relevant characteristics except group variable	Samples were NOT similar for relevant characteristics except group variable	
Statistical methods						
2	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	Not applicable
3	Was the sample size included in the analysis adequate?	Regression analyses or IRT/Rasch based analyses: 200 subjects per group	150 subjects per group	100 subjects per group	< 100 subjects per group	
		MGCFA*: 7 times the number of items and ≥100	5 times the number of items and ≥100; OR 5-7 times the number of items but <100	5 times the number of items but <100	<5 times the number of items	
Other						
4	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

*MGCFA: multi-group confirmatory factor analyses

Box 6. Reliability						
Design requirements		very good	adequate	doubtful	inadequate	NA
1	Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3	Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
Statistical methods						
4	For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	Not applicable
5	For dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			No kappa calculated	Not applicable

6	For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
<i>Other</i>						
8	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 7. Measurement error						
Design requirements		very good	adequate	doubtful	Inadequate	NA
1	Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2	Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3	Were the test conditions similar for the measurements? (e.g. type of administration, environment, instructions)	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
Statistical methods						
4	For continuous scores: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	SEM, SDC, or LoA calculated	Possible to calculate LoA from the data presented		SEM calculated based on Cronbach's alpha, or on SD from another population	Not applicable
5	For dichotomous/nominal/ordinal scores: Was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated		% agreement not calculated	Not applicable
Other						
6	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 8. Criterion validity						
		very good	adequate	doubtful	inadequate	NA
Statistical methods						
1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated			Correlations or AUC NOT calculated	Not applicable
2	For dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated	Not applicable
Other						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 9. Hypotheses testing for construct validity
9a. Comparison with other outcome measurement instruments (convergent validity)
Design requirements

1 Is it clear what the comparator instrument(s) measure(s)?

very good adequate doubtful inadequate NA

Constructs measured by the comparator instrument(s) is clear

Constructs measured by the comparator instrument(s) is not clear

2 Were the measurement properties of the comparator instrument(s) sufficient?

Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population

Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population

Some information on measurement properties of the comparator instrument(s) in any study population

No information on the measurement properties of the comparator instrument(s), OR evidence for insufficient measurement properties of the comparator instrument(s)

Statistical methods

3 Was the statistical method appropriate for the hypotheses to be tested?

Statistical method was appropriate

Assumable that statistical method was appropriate

Statistical method applied NOT optimal

Statistical method applied NOT appropriate

<i>Other</i>					
4	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws

9b. Comparison between subgroups (discriminative or known-groups validity)						
Design requirements		very good	adequate	doubtful	inadequate	NA
5	Was an adequate description provided of important characteristics of the subgroups?	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor or no description of the important characteristics of the subgroups		
Statistical methods						
6	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
7	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws (e.g. only data presented on a comparison with an instrument that measures another construct)	Other important methodological flaws	

Box 10. Responsiveness						
10a. Criterion approach (i.e. comparison to a gold standard)						
		very good	adequate	doubtful	inadequate	NA
Statistical methods						
1	For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated	Not applicable
2	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated	Not applicable
Other						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)						
Design requirements		very good	adequate	doubtful	inadequate	NA
4	Is it clear what the comparator instrument(s) measure(s)?	Constructs measured by the comparator instrument(s) is clear			Constructs measured by the comparator instrument(s) is not clear	
5	Were the measurement properties of the comparator instrument(s) sufficient?	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s) OR evidence of poor quality of comparator instrument(s)	
Statistical methods						
6	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method were appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
7	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)						
Design requirements		very good	adequate	doubtful	inadequate	NA
8	Was an adequate description provided of important characteristics of the subgroups?	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor or no description of the important characteristics of the subgroups		
Statistical methods						
9	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
10	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10d. Construct approach: (i.e. hypotheses testing: before and after intervention)						
Design requirements		very good	adequate	doubtful	inadequate	NA
11	Was an adequate description provided of the intervention given?	Adequate description of the intervention		Poor description of the intervention	NO description of the intervention	
Statistical methods						
12	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
13	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

