

营养干预的循证评价:对现有循证评价方法的系统综述与评价

王波 詹思延 孙凤 李立明

【导读】 探讨建立一个标准、严格的营养干预循证评价方法。通过检索电子数据库、查找参考文献和引文、咨询方法学专家,对现有循证评价方法进行全面综述;通过检索实证研究证据与方法学系统综述、识别广泛接受的方法学对现有循证评价方法进行严格评价,最终纳入12个合格的循证评价方法,识别并解决了9项方法学步骤上的差异,在综合分析的基础上提出了针对营养干预的循证评价方法学标准。本文建立的循证评价方法对营养干预具有良好的适用性。

【关键词】 营养干预;循证评价;系统综述

Evidence-based evaluation of nutritional interventions: a comprehensive review and critical appraisal of current methods WANG Bo¹, ZHAN Si-yan², SUN Feng³, LI Li-ming². 1 Health Science Popularization Research Center, Chinese Academy of Medical Sciences, Beijing 100730, China; 2 Department of Epidemiology and Biostatistics, School of Public Health, Peking University; 3 Department of Epidemiology and Biostatistics, School of Medicine, Shihezi University
Corresponding author: LI Li-ming, Email: lmlec@vip.163.com
This work was supported by a grant from the Doctoral Fund of Ministry of Education of China (No. 20050001159).

【Key words】 Nutritional interventions; Evidence-based evaluation; Systematic review

一、研究背景

尽管当前应用循证方法评价营养干预和膳食疾病关联、制定膳食指南、批准健康要求(health claim)和制定营养政策已经成为共识^[1],但循证评价尚无一个被广泛认可和采用的统一科学证据分析过程^[2],主要体现在2个方面:

1.不同的循证方法在某些方法学步骤上存在差异,并导致完全不同的评价结果。如一项系统综述中评价小组不同的人员构成可能导致评价过程判断的差异^[3],提示人员组成会影响最终的循证评价结果;在个体研究的质量评价方面,一项研究使用25个不同的质量评价工具对17项比较低分子肝素与标准肝素预防术后血栓效果的试验进行质量评价^[4],发现不同的质量评价工具会导致不同的研究质量评价结果,从而导致Meta分析结果的差异;在对证据质量和循证建议进行分级方面,1篇综述识别了57种不同的方法^[5]。

2.营养干预的特殊性:与通常的临床干预相比,对营养干预的评价必须考虑到以下特殊性。①替代结局(surrogate outcome)的使用。考虑到从膳食营养暴露到发生慢性非传染性疾病之间有数十年的时间,许多营养干预研究使用替代结局(如LDL-C、血糖等)来反映干预效果^[6,7]。如果使用的

替代结局其有效性尚未获得证实,则需要对联系营养干预与疾病结局的所有间接证据进行评价^[8]。②公共卫生干预的特点。在公共营养领域,营养干预常常具有公共卫生干预的特点,如评价时应用非随机化试验(由于伦理学或可行性的限制)和群随机化试验^[9,10]、研究异质性(常常由多种研究设计、多成分的干预或多样化的研究对象所造成)^[11],对此类干预的评价在方法学上将面临更多的挑战^[12]。

加拿大健康体检工作组(the Canadian Task Force on the Periodic Health Examination)建议^[13],将现有循证方法进行最佳组合,并进一步优化,从而提供更有意义的建议和更加严格的解释。为此,本研究对现有循证评价方法做全面综述与严格评价,为营养干预的循证评价提供方法学标准。

二、研究方法

1. 系统综述:

(1)纳入/排除标准:纳入的评价方法必须同时符合以下3项标准:①在制定临床指南、公共卫生指南或膳食营养指南(包括膳食指南、健康要求评价、膳食疾病关联评价)时用于评价干预的循证方法;②包含循证评价7项主要内容(建立评价小组、形成问题、检索合格研究、纳入研究和提取数据、评价偏倚危险、整合证据、提出循证建议)中的至少5项;③由国际专业学术机构、专业学术团体、方法学工作组或学术会议所制定。符合以下5项标准中任何一项即被排除:①非循证的指南制定或干预评价方法;②非英文发表或公布的方法;③关于干预效果的系统综述;④对指南或系统综述方法学质量进行评价的方法;⑤仅有循证评价的部分内容(只有

DOI: 10.3760/cma.j.issn.0254-6450.2010.09.024

基金项目:教育部博士点基金项目(20050001159)

作者单位:100730 北京,中国医学科学院健康科普研究中心(王波);北京大学公共卫生学院流行病与卫生统计学系(詹思延、李立明);石河子大学医学院流行病与卫生统计学系(孙凤)

通信作者:李立明, Email: lmlec@vip.163.com

上述 7 项主要内容中的 4 项或更少)。

(2) 检索途径与策略: 与 Cochrane Developmental, Psychosocial and Learning Problems Group 文献检索专家 Joanne Abbott 制定了高灵敏度的检索策略, 检索以下电子数据库: Pubmed, EMBASE, BIOSIS, WHOLIS, National Guideline Clearinghouse (NGC) Annotated Bibliographies 和 Cochrane Library (2008 年第 2 期); 使用 Google 检索 Internet; 检索合格研究的参考文献及引文; 咨询 NGC 的指南方法学专家。检索仅限于英语, 发表年代为 2000—2008 年 9 月。

(3) 方法的选择与信息提取: 先按“题目”和“摘要”筛选有关的研究, 在获取全文的基础上根据纳入/排除标准选择合格的方法。方法入选由 2 名综述者独立完成, 并通过讨论按照“纳入/排除标准”解决不一致。方法学信息提取先由 1 名综述者完成, 再由另 1 名综述者核对。

2. 严格评价: 由 2 名综述者独立进行评价, 并按照严格评价的方法和步骤进行讨论, 解决不一致。评价方法包括 4 个步骤: ①根据各方法总结和综合循证方法的基本框架; ②评价框架内各方法学的一致性; ③检索文献为解决不一致查找证据; ④根据方法学系统综述或实证研究证据为不一致选择方法学标准, 当缺少上述两类证据时以广泛接受的方法学(即关注该方法的方法中, 有一半以上使用的方法学)为标准, 当不存在广泛接受的方法学时, 仅对方法学不一致的现状进行报告。

三、研究结果

1. 纳入的方法: 最终纳入 12 个合格的方法^[14-25](表 1), 其中 9 个方法^[14, 15, 17-19, 21-23, 25]直接关注或涉及对公共卫生干预的循证评价, 4 个方法^[15, 17, 18, 22]直接关注了对营养干预的循证评价。

2. 循证评价的基本框架: 12 个纳入方法具有一致或相近的循证评价步骤。综合后建立了循证评价的基本框架(图 1)。

3. 循证评价的方法: 在系统综述、严格评价的基础上, 经过综合分析, 提出了循证评价基本框架下各主要步骤的方法学标准(表 2)。

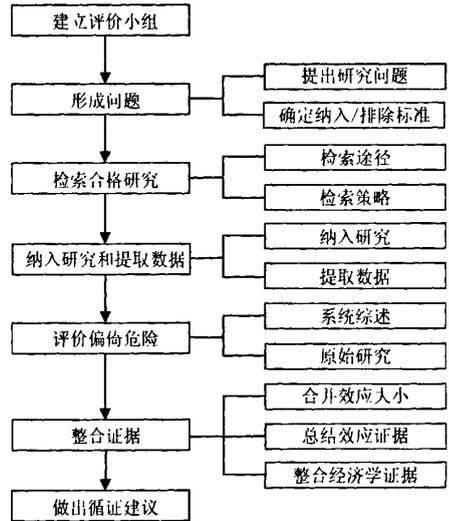


图 1 对干预进行循证评价的基本框架

(1) 建立评价小组: 有 11 个方法^[14-21, 23-25]建立了评价小组, 并要求具备系统综述方法学和所评价主题的相关专业知识技能; 一些方法^[15, 16, 18, 19]的评价小组是由多学科专家组成, 如要求信息检索专家和/或卫生经济学专家参与, 某些临床指南^[16]还有患者代表参与。有 2 个方法^[16, 19]对利益冲突问题提出了明确的政策。

(2) 设计: ①提出研究问题: 有 5 个方法^[14, 16, 19, 24, 25]关注此项。所有方法一致建议使用 PICO 框架, 即研究对象 (participants)、干预 (interventions)、对照 (comparisons) 和结局 (outcomes), 建立格式化明确的研究问题。②确定纳入/排除标准: 10 个方法^[14, 15, 17-19, 21-25]关注了纳入/排除标准的确定, 表现在研究设计的类型、研究对象的类型、干预类型、结局类型 4 个方面。7 个方法^[14, 15, 17, 18, 21, 22, 25]关注研究设计类型的确定, 其中 2 个方法^[21, 25]提出根据干预特征(是否可实现随机分配)、结局特征(是否长期或罕见结局)确定合格研究设计类型的原则, 另外 2 个方法^[14, 18]提出随机分配的干预主要关注

表 1 本研究纳入的 12 个方法学

纳入的方法	制定机构
Cochrane handbook for systematic reviews of interventions (Version 5.0.1) ^[14]	The Cochrane Collaboration
Nutrition, health and related claims: final assessment report ^[15]	Food Standards Australia New Zealand (FSANZ)
Scottish Intercollegiate Guidelines Network: a guideline developer's handbook ^[16]	Scottish Intercollegiate Guidelines Network (SIGN)
Evidence-based review system for the scientific evaluation of health claims ^[17]	Food and Drug Administration (FDA)
Food, nutrition, physical activity and the prevention of cancer: a global perspective ^[18]	The World Cancer Research Fund (WCRF) & the American Institute for Cancer Research (AICR)
National Institute for Health and Clinical Excellence: the guideline manual and methods for development of NICE public health guidance ^[19]	National Institute for Health and Clinical Excellence (NICE)
The guide to clinical preventive services-methods ^[20]	The U.S. Preventive Services Task Force (USPSTF)
Improving the use of research evidence in guideline development ^[21]	WHO Advisory Committee on Health Research
Process for the assessment of scientific support for claims on foods(PASSCLAIM): Consensus on criteria ^[22]	ILSI Europe Functional Food Task Force
The guide to community preventive services-methodological background ^[23]	The Task Force on Community Preventive Services (TFCPS)
Handbook for the preparation of explicit evidence-based clinical practice guidelines ^[24]	New Zealand Guidelines Group (NZGG)
Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews ^[25]	NHS Centre for Reviews and Dissemination (CRD)

表2 循证评价基本框架中各主要步骤的方法学标准

主要步骤	方法学标准
建立评价小组	建立具有系统综述方法学、对所评价主题相关专业知识的多学科评价小组,并有信息检索专家和卫生经济学专家参与;所有成员必须公开潜在的利益冲突
形成问题	
提出研究问题	使用PICO框架建立格式化的研究问题
确定纳入/排除标准	研究设计主要关注随机化试验,当下难以实现随机分配或有长期、罕见结局时,才考虑纳入非随机干预试验、队列研究和病例对照研究;只有在研究的终点结局难以直接测量,才应根据生物学有效性、相关性和方法学有效性谨慎选择替代结局,并根据评价目的决定是否采用有害结局和经济学结局
检索合格研究	
检索途径	Medline、EMBASE和Cochrane Library等数据库,还应包括手工检索、参考文献及引文检索、专家咨询等
检索策略	在信息检索专家指导下制定全面合理的检索策略;必须对检索策略和检索结果进行记录、报告
纳入研究和提取数据	
纳入研究	由2人独立完成,讨论解决分歧
提取数据	由2人独立完成,讨论解决分歧
评价偏倚危险	分别使用SIGN、EPOC和SIGN清单评价随机对照试验、非随机干预试验和观察性研究(队列研究和病例对照研究)的方法学质量,使用Drummond清单评价卫生经济学研究方法学的质量;质量评价由2人独立完成,并讨论解决分歧
整合证据	
合并效应大小	合并方法包括叙述性合成与Meta分析;通过对异质性的定性与定量评价选择合适的合并方法;使用亚组分析或Meta回归方法调查异质性的来源
总结效应证据	使用GRADE系统对证据总体质量进行分级;根据USPSTF标准判断间接证据的不确定性程度
整合经济学证据	使用叙述性合成对经济学证据(尤其是证据质量)进行总结(适时进一步使用定量方法)
做出循证建议	使用GRADE系统对证据循证建议进行分级

随机化试验(包括群随机对照试验)。而对于由于伦理学或可行性的原因难以实现随机分配的干预,需要评价的结局为罕见或长期结局(尤其是有害效应结局)则考虑纳入非随机干预试验(有对照的前后比较研究和中断时间序列)、队列研究和病例对照研究。8个方法^[14,15,17,21-25]关注研究结局类型。3个方法^[14,17,24]强调要谨慎选择替代结局(当终点结局难以直接测量时,可以考虑使用替代结局)。另外2个方法^[15,22]则提出替代结局必须具有生物学有效性(替代结局与终点结局之间应有因果关联)、生物学相关性(替代结局应位于暴露与终点结局之间)和方法学有效性(替代结局能够获得可信、精确的测量)。除了关注预期的效应外,4个方法^[15,21,23,25]提出还应根据评价目的,考虑关注有害效应结局和/或经济学结局。

(3)检索:①检索途径:9个方法^[14-16,18-20,22-25]描述了常用的检索途径,包括电子数据库(包括基本数据库和专题数据库)、手工检索(包括未被电子数据库收录的期刊和会议文摘)、参考文献及引文检索、专家咨询和试验注册库。基本数据库至少应包括Medline、EMBASE和Cochrane Library。②检索策略:11个方法^[14-20,22-25]关注了与检索策略有关的问题。所有方法均强调应制定全面的检索策略,尽最大可能检索到所有相关的文献;7个方法^[14-16,18,19,24,25]明确要求在信息检索专家的指导或参与下制定检索策略;2个方法^[14,18]指出检索策略不应受语言、检索时间段(除非相关研究只发表在特定期限)进行限制;5个方法^[14,18,19,24,25]提出必须对检索策略和检索结果进行完整记录,并在结果中进行报告。

(4)纳入研究和提取数据:①纳入研究:8个方法^[14-16,18-20,24,25]关注了与研究纳入有关的问题。其中6个方法^[14,15,18,20,24,25]要求纳入过程由2人独立完成,并通过讨论解决不一致(基于纳入/排除标准),必要时(2人讨论无法达成一致)通过全组讨论或参照第3人意见解决;2个方法^[16,19]提出研究纳入过程

可由1人完成,并通过全组讨论解决不一致。一项研究证实^[26],由2人独立完成可降低纳入研究中的遗漏。②提取数据:5个方法^[14,18,20,23,25]关注与提取数据有关的问题。其中4个方法^[14,18,23,25]要求应由2人独立完成数据提取,并以讨论形式解决不一致(基于数据提取表格与要求),必要时(讨论无法达成一致)通过全组讨论或参照第三者意见解决;1个方法^[20]提出数据的提取可由1人完成。一项实证研究发现^[27],由2人独立完成数据提取,相比由1人完成并经另1人确认的方法,更少出现错误。

(5)评价偏倚危险:所有方法都强调了对个体研究的方法学质量评价,6个方法^[16,19,20,23-25]明确要求质量评价应由2人独立完成,并通过讨论解决不一致(基于质量评价方法与工具)。9个方法^[14-17,19,20,22-25]具体关注了相应的评价工具,其中2个方法^[15,23]使用单一的质量评价清单对不同类型的研究设计进行评价。然而,对不同研究设计进行质量评价已成为目前广泛接受的方法学标准(质量评价的方法与关注质量评价工具的方法数量之比为7:9)。

7个方法^[14,16,17,19,20,24,25]关注了随机对照试验的质量评价。其中2个方法^[16,19]使用了苏格兰学院间指南网络(Scottish Intercollegiate Guideline Network, SIGN)清单;另外5个方法^[14,17,20,24,25]分别使用Cochrane自制清单、FDA自制清单、美国预防服务工作组(U.S. Preventive Services Task Force, USPSTF)自制清单、A Generic Appraisal Tool for Epidemiology(GATE)清单和Verhagen清单。根据美国医疗保健研究与质量局(Agency for Healthcare Research and Quality, AHRQ)的“最佳实践”指南^[28],该工具应包括对以下5个与内部有效性有关的重要领域:随机化、盲法、结局、统计分析和资金支持。SIGN清单包括了所有的5个重要领域;FDA自制清单、USPSTF自制清单和GATE清单缺乏对资金

资助的评价;Cochrane 自制清单缺乏对统计分析和资金资助 2 个领域的评价;Verhagen 清单缺乏对结局和资金资助 2 个领域的评价。

2 个方法^[14,19]关注了用于非随机干预试验的质量评价工具,均使用 Cochrane Effective Practice and Organisation of Care Group (EPOC)清单评价有对照的前后比较研究和中断时间序列。7 个方法^[14,16,17,19,20,24,25]关注了用于观察性研究(队列研究和病例对照研究)的质量评价工具。除 1 个方法^[17]使用统一的工具(FDA 自制清单)对队列研究和病例对照研究进行评价外,另外 6 个方法^[14,16,19,20,24,25]均使用不同工具对队列研究和病例对照研究进行评价;3 个方法^[14,20,25]分别使用 Newcastle-Ottawa Scale (NOS)量表、USPSTF 自制清单和 Centre for Reviews and Dissemination (CRD)自制清单对队列研究和病例对照研究进行评价;2 个方法^[16,19]均使用 SIGN 清单分别对队列研究和病例对照研究进行评价;1 个方法^[24]使用 GATE 清单对队列研究进行评价。根据 AHRQ 的“最佳实践”指南^[28],观察性研究的评价工具应包括对以下 5 个重要领域:研究对象可比性、暴露、结局测量、统计分析和资金资助。

5 个方法^[14,16,19,23,25]关注了用于经济学评价研究的质量评价工具。其中 3 个方法^[14,19,25]推荐使用 Drummond 清单;1 个方法^[14]推荐 The Consensus Health Economic Criteria (CHEC)清单;另外 2 个方法^[16,23]分别使用了 SIGN 清单(基于 MERGE 基础上制定)和 TFCPS 自制清单。可见,Drummond 清单是目前广泛接受的方法学标准(使用该清单的方法与所有关注经济学评价研究质量评价工具的方法之数量比为 3:5)。

(6)整合证据:①合并效应:4 个方法^[14,18,23,25]关注了与合并方法有关的问题。所有方法都将合并方法分为叙述性合成(narrative synthesis)和定量合成两类;3 个方法^[14,18,25]关注了叙述性合成的方法,即通过表格、森林图对研究特征(如设计、对象、结局、质量等)与研究结果进行比较和总结,定性评价研究结果在不同研究特征上是否相似(即研究结果是否与某些研究特征有关);3 个方法^[14,18,25]指出当研究之间存在重要异质性时不适合使用 Meta 分析。1 个方法^[14]提出,使用 Meta 分析对观察性研究进行结果合并时,如存在调整混杂因素的研究,则应使用广义倒方差法进行合并。该方法^[14]还提出了整合群随机对照试验的方法:当群随机化试验不存在分析单位错误时,直接使用广义倒方差法进行效应合并;当存在分析单位错误时,首先使用相似研究提供的群内相关系数(Intracluster correlation coefficient)、平均群大小计算设计效应,再根据设计效应进行近似调整分析后使用广义倒方差法对效应进行合并(同时应对这种近似调整进行敏感性分析)。5 个方法^[14,18,19,23,25]关注了与异质性调查和处理有关的问题,并提出基本一致的方法学(图 2):首先对纳入研究的异质性进行定性评价,判断是否存在重要异质性;如存在重要异质性则根据异质性程度使用叙述性合成或亚组 Meta 分析,如不存在重要异质性则根据异质性检验结果选择固定或随机效应模型进行 Meta 分析;不论异质性定量评价结果如何,均使用亚组分析或 Meta 回归的方法调查异质性来源。3 个方法^[14,18,25]进一步关注了调查异质性来源时应注意的问

题。在异质性调查变量的选择上,1 个方法^[11]要求应在科学依据基础上预先确定调查变量,且变量数目应尽可能地少。在异质性调查结果的解释上,3 个方法均要求谨慎对待异质性调查结果。②总结效应证据:12 个方法^[14-25]均关注了对证据总体质量的分级方法。3 个方法^[14,19,21]使用 The Grading of Recommendations Assessment, Development and Evaluation (GRADE)系统;9 个方法^[15-18,20,22-25]分别使用了自制的分级系统。GRADE 系统(表 3)是由 1 个国际指南方法学专家组在对已有的 57 个证据质量分级系统和循证建议分级系统进行系统综述和严格评价的基础上制定的^[5,29,30]。此外,有 1 个方法^[20]提出了判断间接证据不确定性的标准:即证据的完整性、各关联的拟合程度、干预与健康结局之间证据直接性的程度(即关联的数量)。③整合经济学证据:5 个方法^[14,19,21,23,25]关注了评价干预时对经济学证据的整合。其中 1 个方法^[21]指出,是否需要整合经济学证据需要根据所关注的研究问题来确定;4 个方法^[14,21,23,25]均使用叙述性合成对经济学评价研究的特征与结果进行总结;2 个方法^[14,21]提出可适当考虑使用定量方法(Meta 分析或建立模型);2 个方法^[14,19]均强调在整合经济学证据时,必须重视对其质量的评价。

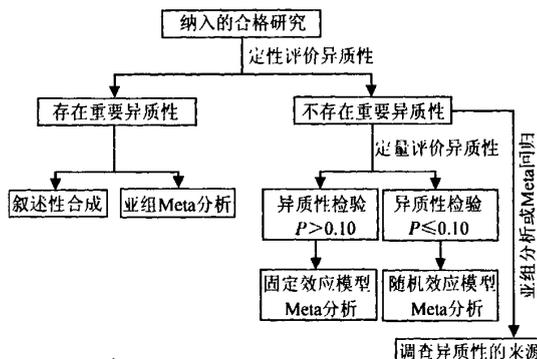


图 2 异质性的调查与处理方法

表 3 GRADE 系统对证据总体质量的分级

质量分级	分级标准 ^a
高	随机化试验,或质量升高 2 级的观察性研究
中	质量降低 1 级的随机化试验,或质量升高 1 级的观察性研究
低	质量降低 2 级的随机化试验,或观察性研究
很低	其他类型的证据,或质量降低 3 级的随机化试验,或质量降低 1 级的观察性研究

注:^a有 5 个降低证据质量的因素:严重偏倚;存在不确定性的非直接证据;研究结果不一致;结果不精确;可能存在发表偏倚。当存在上述任何一个因素时,证据质量即降低 1 级(最多降低 3 级);当研究质量或证据存在非常严重的局限性时,证据质量应降低 2 级。有 3 个升高证据质量的因素:大的干预效应(RR 值 > 2 或 < 0.5);存在剂量反应关系;潜在的混杂降低已经证实的干预效应。当存在任何一个因素时,证据质量即升高 1 级(最多升高 3 级);当干预效应非常大时(RR 值 > 5 或 < 0.2),证据质量升高 2 级

(7)做出循证建议:11 个方法^[15-25]关注了如何制定循证建议。其中 1 个方法^[21]使用了基于对已有 57 个方法的系统综述和严格评价^[5,29,30]制定的 GRADE 系统循证建议分级方法(图 3,表 4),即基于证据总体质量和利弊权衡,考虑资源使用(经济学证据)和价值观选择,将循证建议分为高、低两级;其他 10 个方法^[15-20,22-25]均使用了自制的循证建议分级系统。

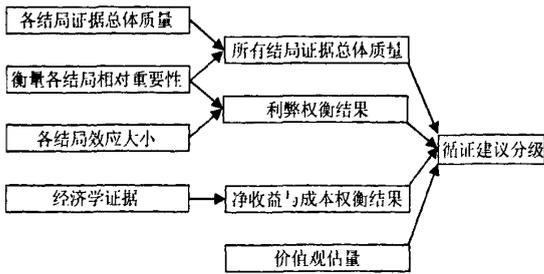


图3 GRADE 系统制定循证建议的过程

表4 GRADE 系统对循证建议分级时考虑的因素

因素	该因素对分级的影响
利弊权衡	利弊之间的差异越大,高度的建议可能越大
证据质量	证据质量越高,高度的建议可能越大
价值观	价值观差异越大,或不确定性越大,低度的建议可能越大
成本资源	干预成本过低或过高时,高度的建议可能越大

四、结论

1. 关于研究方法:本研究并未对所有相关的方法学研究进行系统综述,而是采纳加拿大健康体检工作组的建议^[13],对现有循证评价方法做总结与评价,以建立一个新的、更加严格的营养干预循证评价统一的方法。主要基于以下考虑:①检索策略和工作量的限制。对方法学研究的检索,通常很难设计一个足够灵敏、有效的检索策略来识别合格的研究^[31],且系统综述通常会检索出数目庞大的文献^[32]。②通常在制定循证方法时都会在一定程度上考虑和整合了最新的实证研究证据和方法学共识^[14-25],因而即使开展方法学系统综述,也不大可能获得超出本研究方法学标准之外的实证信息^[33]。

在方法学的综述中,纳入了8个^[14,16,19-21,23-25]未直接关注营养干预的循证评价方法。这些方法虽然未直接关注营养干预,却关注了更为广泛的“干预”(包括营养干预)。由于这些方法未注意到营养干预的特殊性(包括公共卫生干预的特殊性),其循证评价方法学的质量必然不如本研究所建立的方法。

本研究发现循证方法中有9项不一致,并从现有证据(方法学系统综述或实证研究)或广泛接受的方法学中选择方法学的标准。与既往的方法学系统综述^[21,28]不同,以“广泛接受的方法学”作为标准时,其定义是在关注的问题中使用的方法学有一半以上为该方法。

对合格方法的检索,本研究只收录英文文献中的循证方法,基于如下考虑:①循证方法的英文文献占主流;②与干预效果评价不同,对方法学研究的综述与评价更强调重要文献的检索^[32];③纳入其他语种的文献会增加研究难度和成本^[34]。

2. 关于研究结果:本研究建立的方法是基于对12个循证方法严格评价之上。这些方法均由国际权威学术机构制定且多数已获得广泛的应用,其中包括对公共卫生干预和营养干预进行循证评价的方法,因此本研究所建立的循证评价方法具有公共卫生干预和营养干预的特点,增强了研究结果对营养干预的适用性。此外,本研究所建立的循证评价方法

涵盖了系统综述、证据质量分级、循证建议分级等过程,并未涉及指南形成的过程。

本研究特别关注替代结局的使用、间接证据的评价、非随机化试验和群随机化试验的应用、异质性研究,此外营养干预的评价还受另两类问题的影响:营养干预研究实施的质量和膳食营养干预评价结果的解释。第一类问题包括膳食营养暴露测量方法的准确性、潜在混杂因素测量方法的准确性和对潜在混杂因素的处理等。近年来,在膳食疾病关联研究领域中对膳食营养暴露调查方法(包括24 h膳食回顾、手工膳食记录、膳食记录、食物频率问卷等)的选择颇有争议^[35],且无一种方法能完全真实地反映个体的膳食摄入状况^[36](如能量摄入都会被低估20%~30%^[37]);检测尿氮水平以估计蛋白质摄入情况,男性总会过高报告蛋白质摄入量,而女性则准确^[38]。测量混杂因素与膳食营养暴露同样复杂,研究表明混杂因素的偏倚即使经过统计学调整,仍会歪曲膳食疾病真实的关联^[39]。此类问题在对非膳食营养干预评价中同样存在(在膳食疾病关联研究领域更为突出^[40]),本研究建立的循证评价方法已注意到对原始研究的偏倚危险评价。第二类问题包括营养素基线暴露、营养素生物利用率、营养素之间的多重共线性等,但此类问题并不影响循证评价的过程^[6]。

对个体研究的方法学质量评价,加拿大卫生技术评估协作办公室(the Canadian Coordinating Office for Health Technology Assessment)使用相同的方法(即基于AHRQ“最佳实践”指南)对2000—2005年发表的230个质量评价工具进行系统综述与严格评价,获得了与本研究完全一致的结果^[41]。

GRADE系统中的证据质量分级和循证建议分级已被全球20多个学术机构或组织采用^[14,42-44]。然而并非完整,如缺乏对非随机化干预试验(有对照的前后比较研究、中断时间序列)的关注、未区分队列研究与病例对照研究在证据质量上的不同。随着方法学的进展,GRADE系统也将不断获得改进^[45]。在对现有循证方法的证据质量分级系统进行严格评价时,发现USPSTF分级系统^[20]所提出的用于判断间接证据不确定性的标准,可以与GRADE证据分级系统对接,从而更明确、客观地定义间接证据的不确定性。

经济学证据能够提高公共卫生决策的效率,是循证决策的重要依据之一^[19]。然而与干预研究结果的合成方法相比,经济学评价的合成方法还有待进一步发展^[19,46]。

(感谢英国Bristol大学Joanne Abbott对文献检索提供的帮助)

参 考 文 献

- [1] Anderson GH, Black R, Harris S. Dietary guidelines: past experience and new approaches. *J Am Diet Assoc*, 2003, 103 Suppl:S3-4.
- [2] Myers E. Systems for evaluating nutrition research for nutrition care guidelines: Do they apply to population dietary guidelines? *J Am Diet Assoc*, 2003, 103 Suppl:S34-41.
- [3] Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*, 1998, 2:i-88.
- [4] Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*, 1999, 282: 1054-1060.
- [5] Atkins D, Eccles M, Flottorp S, et al. GRADE Working Group.

- Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches. *BMC Health Serv Res*, 2004, 4: 38-44.
- [6] Freudenheim JL. Study design and hypothesis testing: issues in the evaluation of evidence from research in nutritional epidemiology. *Am J Clin Nutr*, 1999, 6 Suppl: S1315-1321.
- [7] Truswell AS. Levels of evidence for nutritional science: a workshop report. *Food Aust*, 2001, 53: 272-276.
- [8] Lupton JR. Determining the strength of the relationship between a food, food component, or dietary supplement ingredient and reduced risk of a disease or health-related condition. *J Nutr*, 2005, 135: 340-342.
- [9] Sempos CT, Liu K, Ernst ND. Food and nutrient exposures; what to consider when evaluating epidemiologic evidence. *Am J Clin Nutr*, 1999, 69 Suppl: S1330-1338.
- [10] Margetts B, Warm D, Yngve A, et al. Developing an evidence-based approach to public health nutrition: translating evidence into policy. *Public Health Nutr*, 2001, 4: 1393-1397.
- [11] Jackson SF, Edwards RK, Kahan B, et al. An assessment of the methods and concepts used to synthesize the evidence of effectiveness in health promotion: a review of 17 initiatives. Canadian Consortium for Health Promotion Research, Toronto, 2001.
- [12] Jackson N, Waters E. For the Guidelines for Systematic Reviews in Health Promotion and Public Health Taskforce. Criteria for the systematic review of health promotion and public health interventions. *Health Promot Int*, 2005, 20: 367-374.
- [13] Canadian Task Force on Preventive Health Care (CTFPHC). Methodology. <http://www.ctfphc.org/resources&pubs.htm>.
- [14] Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]*. The Cochrane Collaboration, 2008.
- [15] Food Standards Australia New Zealand. Proposal P293: nutrition, health and related claims (final assessment report). 11 April 2008. www.foodstandards.gov.au/foodmatters/health_nutrition_and_relatedclaims/index.cfm.
- [16] SIGN 50: A guideline developers' handbook (revised edition). January 2008. www.sign.ac.uk/guidelines/fulltext/50.
- [17] Food and Drug Administration. Guidance for industry: evidence-based review system for the scientific evaluation of health claims. July 2007. www.cfsan.fda.gov/~dms/hclmguis.html.
- [18] Marmot M, Atinmo T, Byers T, et al. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. WCRF/AICR Expert Report, 2007.
- [19] National Institute for Health and Clinical Excellence. The guidelines manual. April 2007. www.nice.org.uk/guidelinesmanual.
- [20] Harris RP, Helfand M, Woolf SH, et al. Current methods of the U.S. Preventive Services Task Force: a review of the process. *Am J Prev Med*, 2001, 20: 21-35.
- [21] Norwegian Knowledge Centre for the Health Services. Improving the use of research evidence in guideline development. Health Research Policy and Systems, 2006.
- [22] Aggett PJ, Antoine JM, Asp NG, et al. Process for the assessment of scientific support for claims on foods (PASSCLAIM): consensus on criteria. *Eur J Nutr*, 2005, 44: 1/1-1/30. DOI 10.1007/s00394-005-1101-6.
- [23] Briss PA, Zaza S, Pappaioanou M, et al. The task force on Community preventive services. Developing an evidence-based guide to community preventive services—methods. *Am J Prev Med*, 2000, 18: 35-43.
- [24] New Zealand Guidelines Group. Handbook for the preparation of explicit evidence-based clinical practice guidelines. New Zealand, 2001.
- [25] The Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. 2001. www.york.ac.uk/inst/crd/pdf/crdreport4_complete.pdf.
- [26] Edwards P, Clarke M, DiGiuseppi C, et al. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Stat Med*, 2002, 21: 1635-1640.
- [27] Buscemi N, Hartling L, Vandorneer B, et al. Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol*, 2006, 59: 697-703.
- [28] West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evidence report/technology assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality, 2002.
- [29] Atkins D, Briss PA, Eccles M, et al. GRADE Working Group. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res*, 2005, 5: 25-36.
- [30] Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, 2008, 336: 924-926.
- [31] The Cochrane Methodology Register (CMR) Issue 1 2007. http://www3.cochrane.org/access_data/cmhr/help.htm. Accessed November 5, 2008.
- [32] Ukoumunne OC, Gulliford MC, Chinn S, et al. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess*, 1999, 3: iii-92.
- [33] Schünemann HJ, Frertheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Health Res Policy Syst*, 2006, 4: 13 doi: 10.1186/1478-4505-4-13.
- [34] Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *Chalmers I, Altman DG. Systematic reviews*. London: BMJ, 1995: 17-36.
- [35] Sempos CT, Briefel RR, Flegal KM, et al. Factors involved in selecting a dietary survey methodology for national nutrition surveys. *Aust J Nutr Diet*, 1992, 49: 96-101.
- [36] Briefel RR, Sempos CT, McDowell MA, et al. Dietary methods research in the third national health and nutrition examination survey: underreporting of energy intake. *Am J Clin Nutr*, 1997, 65: 1203S-1209S.
- [37] Martin LJ, Su W, Jones PJ, et al. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary-intervention trial. *Am J Clin Nutr*, 1996, 63: 483-490.
- [38] Kahn HA, Whelton PK, Appel LJ, et al. Validity of 24-hour dietary recall interviews conducted among volunteers in an adult working community. *Ann Epidemiol*, 1995, 5: 484-489.
- [39] Marshall J, Hastrup JL. Mismeasurement and the resonance of strong confounders: uncorrelated errors. *Am J Epidemiol*, 1996, 143: 1069-1078.
- [40] Margetts BM, Nelson M. Introduction/Margetts BM, Nelson M. Design concepts in nutritional epidemiology. New York: Oxford University Press, 1991: 1-9.
- [41] The Canadian Coordinating Office for Health Technology Assessment. Proposed evaluation tools for COMPUS: opportunity for input. November, 2005. www.cadth.ca/media/pdf/COMPUS_Evaluation_Methodology_draft_e.pdf.
- [42] Schünemann HJ, Jaeschke R, Cook DJ, et al. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med*, 2006, 174: 605-614.
- [43] Guyatt G, Gutterman D, Baumann MH, et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force. *Chest*, 2006, 129: 174-181.
- [44] Guyatt G, Vist G, Falck-Ytter Y, et al. An emerging consensus on grading recommendations? *ACP J Club*, 2006, 144: A8-9.
- [45] The Grading of Recommendations Assessment, Development and Evaluation Working Group. www.gradeworkinggroup.org/index.htm.
- [46] Cochrane Economics Methods Group. Cochrane Economic Methods Group. The Cochrane Library [cdrom]. Issue 3: Oxford: Update Software, 2000.

(收稿日期: 2010-04-16)

(本文编辑: 张林东)