

• 学术争鸣 •

我国护理测量工具文献中内容效度指数应用误区分析

张晨,周云仙

Analysis on errors regarding content validity index used in evaluation of measurement tools in Chinese nursing literature Zhang Chen, Zhou Yunxian

摘要:目的 分析我国护理文献中应用内容效度指数进行测量工具内容效度评价时存在的主要问题,帮助护理研究者更加科学、有效地开发测量工具。方法 检索中国期刊全文数据库(CNKI)、万方数据库及中国生物医学文献数据库(CBM),纳入收录的测量工具开发中涉及内容效度的护理期刊文献,由2名研究者根据评价标准汇总,指出其中的错误及不当做法,并提出参考性建议。结果 共纳入110篇文献,在专家选择方面,10.00%文献未从足够丰富的专业或领域选择专家,86.36%文献对专家所处地区描述不详,84.55%对专家的测量工具构建能力描述不详;在函询方法运用与描述方面,91.82%文献未描述咨询函内容,87.27%未评价测量工具的全面性;CVI计算与呈现方面,94.55%未先计算评定者间一致性,63.64%未标识S-CVI的类型。结论 内容效度指数在护理测量工具开发领域应用广泛,但存在一些认识与应用上的误区,主要在专家选择与描述、专家函询与评分、内容效度指数的计算与呈现三方面。应提高护理研究者对内容效度指数相关知识的认识,掌握正确的应用方法,减少对内容效度指数的不规范应用。

关键词:内容效度; 内容效度指数; 测量工具; 工具开发; 专家选择; 专家函询; 护理科研; 评定者间一致性
中图分类号:R47; G526.3 **文献标识码:**A **DOI:**10.3870/j.issn.1001-4152.2020.04.086

随着我国护理科研的发展,测量工具的开发及汉化日益广泛。这其中常涉及信效度检验。其中效度是指某一测量工具能真正反映其所要测量概念的程度,主要用内容效度、结构效度和效标效度等指标来反映^[1]。内容效度(Content Validity)是指测量工具中的条目能够充分展现所测量内容的程度^[1],是测量工具质量的重要体现^[2]。有学者^[3]认为,提高一项测量工具的内容效度是改善其结构效度的早期关键步骤。内容效度评价主要通过专家函询来获得条目与测量内容的相关信息,并计算相应的指标。目前,在护理研究者中,对测量工具内容效度进行量化最常用的指标是内容效度指数(Content Validity Index,CVI)^[2,4]。有学者将CVI与其他指标进行了比较,认为其具有计算简单、易于理解、注重相关性而非专家间的一致性以及能同时提供条目和测量工具评价信息等优点^[3-4]。然而,在实际护理研究报告中,常常存在对CVI认识及应用上的误区。本文主要对我国护理文献测量工具开发中应用CVI进行内容效度检验时存在的主要问题进行分析,并提出针对性的参考建议,以期减少相应的不规范应用,帮助护理研究者更加科学、有效地开发测量工具。

1 资料与方法

1.1 一般资料 以中国期刊全文数据库(CNKI)、万方数据库及中国生物医学文献数据库(CBM)为检索源。检索内容为护理领域涉及内容效度的文献。检

索策略为(量表 OR 问卷)AND(信效度 OR 效度)。纳入标准:①我国护理领域公开发表的测量工具开发方面涉及内容效度的文献;②发表时间范围为2018年7月1日至2019年6月30日;③发表期刊为2018年中国科技期刊引证报告中纳入核心期刊库的10种护理类期刊,分别是《中华护理杂志》、《护理学杂志》、《中国护理管理》、《中国实用护理杂志》、《护理学报》、《护理研究》、《解放军护理杂志》、《现代临床护理》、《中华现代护理杂志》和《护士进修杂志》。排除标准:①未应用CVI进行内容效度检验的文献;②重复的文献。

1.2 检索结果 分别于CNKI、万方数据库及CBM中检索得到文献89篇、161篇及95篇。根据纳排标准剔除文献235篇,其中不符合纳入标准的文献28篇,未应用CVI进行内容效度检验的文献33篇,重复文献174篇。最终纳入110篇文献进行评析。

1.3 论文评析过程 根据内容效度内涵、测量工具内容效度评价相关权威文献^[1-8]制定评价标准,并提交1名公共卫生、1名统计学及1名量表研制专家审核通过。为减少偏倚,由2名研究者分别根据评价标准分析现有文献,指出其中的错误认识及不当做法,并提出参考性建议。若2名研究者的意见不完全相同,则在双方讨论基础上请第3名测量工具研制专家独立提出意见。

2 结果

在我国护理领域相关文献中,CVI的应用主要在专家选择与描述、专家函询与评分、CVI的计算与描述这三方面存在问题。

2.1 专家选择与描述方面存在的问题 见表1。

作者单位:浙江中医药大学(浙江 杭州,310053)

张晨:女,硕士在读,学生

通信作者:周云仙,1193138211@qq.com

科研项目:浙江中医药大学国家自然科学基金项目(2019ZG41)

收稿:2019-09-23;修回:2019-11-13

表 1 专家选择与描述方面存在的问题(n=110)

具体表现	数量	百分率(%)
专家选择不当		
未从足够丰富的专业或领域选择	11	10.00
未从多个地区选择	5	4.55
专家信息描述不当		
所处地区描述不详	95	86.36
测量工具构建能力描述不详	93	84.55
学历、职称描述不详	24	21.82
数量描述不详	6	5.45

2.2 专家函询与评分方面存在的问题 见表 2。

表 2 专家函询与评分方面存在的问题(n=110)

具体表现	数量	百分率(%)
函询方法运用与描述不当		
未描述咨询函内容	101	91.82
未评价测量工具的全面性	96	87.27
评分方法运用与描述不当		
未描述评分方法的类型	23	20.91
误用德尔菲法的条目重要性评价计算 CVI	3	2.73

2.3 CVI 的计算与呈现方面存在的问题 见表 3。

表 3 CVI 的计算与呈现方面存在的问题(n=110)

具体表现	数量	百分率(%)
计算 CVI 的步骤和方法不当		
未先计算评定者间一致性	104	94.55
未正确计算 CVI	3	2.73
CVI 计算方法和结果呈现不当		
未描述计算 I-CVI 的方法	5	4.55
未报道 I-CVI 的计算结果	10	9.10
未标识 S-CVI 的类型	70	63.64
未根据 CVI 计算结果正确评价内容效度	17	15.45

3 讨论

本文通过对 110 篇相关文献进行分析整理、归纳出护理领域测量工具 CVI 应用中存在的 3 大问题。其中,未阐明 CVI 的计算方法、未说明 S-CVI 的类型、未报道 I-CVI 的计算结果等问题与 Polit 等^[5]的研究结果一致。笔者针对所归纳的问题提出了相应的建议,具体如下。

3.1 选择合适的专家并详细描述 目前没有客观的测量方法来检验测量工具的内容效度。研究人员通常依靠专家小组对测量工具进行相关性评价来计算 CVI^[3]。因此,专家的选择在内容效度评价中十分重要。专家小组成员的信息是帮助读者更好地判断该测量工具内容效度可靠性的重要因素,因此研究者在论文中详细介绍专家的区域、职务、资历等方面的信息十分重要。

本研究发现,所纳入的 110 篇文献中,部分研究未能体现出专家所处专业或领域的丰富性,如均为护

理临床工作者或护理专业教师等,而无临床医生、心理学专家或医院管理者等其他与研究内容相关的专业人员;部分研究未从多个地区选择进行 CVI 评价的专家成员,可能会局限测量工具的使用范围。在专家信息的描述上,部分研究未能详细介绍专家的数量、学历、职称、测量工具构建能力和所处地区等,从而无法判断专家进行 CVI 评价的资质和 CVI 评价的有效性。

内容效度的评价建立在大量文献查阅、工作经验以及综合分析、判断的基础之上^[1]。因此,建议选择熟悉测量内容,具有相关资质并了解目标人群的专家小组成员,且研究者在选择专家时应在专业和学科方面进行良好地组合,如专家小组成员既包括临床医生、护士、教师和研究人员,也包括有丰富的工具构建经验的专家^[6]。此外,如果测量工具将进行广泛应用,因为语言等可能存在区域差异,建议邀请不同国家或同一国家不同区域的专家^[3]。在专家人数方面,由于一般建议进行两轮评价,因此不同阶段的专家选择也有不同。在初始评价阶段,建议选取由 8~12 名成员组成的专家小组来修订或删除有问题的条目,或增加新条目,以充分涵盖有关范畴。研究者依照专家意见对测量工具进行修改后再进行第二轮评价。在第二轮评价中,可选择 3~10 名专家小组成员,其目的是正式评价条目及测量工具内容的有效性^[3,6]。在该轮评价中,建议选用来自首次评价小组的部分专家,因为这样可以依据首次专家评价的信息来选择最合适的评价者。例如,可以基于第一轮评价结果识别出那些不理解或不熟悉评价内容的专家,那些倾向于给高或低评级的专家,或者那些可能有偏见的专家^[3]。2 次评价的时间最好间隔 10~14 d,以免由于时间过短,专家对第 1 次评价结果尚有印象,而影响第 2 次的评价结果^[8]。

3.2 选用恰当的咨询函和 CVI 评价表 专家函询是内容效度评价的主要组成部分,其能影响最终评价结果。其中,应用 CVI 评价表进行评分是专家函询的重要环节。明确、恰当的评分表和评分标准是研究严谨性和可靠性的重要保证。可能由于对咨询函的重视不足,仅少量研究描述了进行专家函询时所采用的咨询函的基本内容;仅少量研究在请专家进行内容效度评价的同时通过邀请专家补充条目等方式来评价测量工具的全面性。在评分方面,部分研究未描述进行内容效度评价时专家所采用的评分方法类型;有研究在进行内容效度评价时,误用德尔菲法的条目重要性评价来计算 CVI。

介绍发给专家的咨询函,能够帮助读者了解专家对该项测量工具评价的了解程度,增加可靠性。建议专家咨询函包括介绍信、研究简介、CVI 评价表和意见征求表^[7,9]。此外,还可以向专家小组简要介绍相

关文献和参考书目^[3]。介绍信建议说明开发该测量工具的原因及意义,选择该名专家的原因及需要专家评价的内容。研究简介建议说明该项测量工具的主要内容、相关概念的定义和概念所包含的维度等^[6],从而让专家了解理论背景以更好地提供评价。*CVI*的缺点在于其只关注测量工具已有条目与相应内容的相关性,对测量工具所包含的条目的综合性和全面性的关注不足^[5]。因此,建议在意见征求表中设置关于条目综合性和全面性的问题,如增设“增加条目”栏,询问专家条目是否有明显的遗漏,测量工具中的每个维度是否有足够数量的条目等^[2-3]。在*CVI*评价表中,专家需对每一个条目进行评价。因奇数等级评分制常出现矛盾的中间等级^[8],一般推荐采用4分制,其中“1”代表该条目与研究内容完全不相关;“2”代表该条目与研究内容弱相关,若采用需要进行重大修订;“3”代表该条目与研究内容比较相关,若采用需要小的修改;“4”代表该条目与研究内容非常相关^[1]。若评分标准不明确,则读者较难判断内容效度评价的有效性。另外,有研究采用德尔菲法的内容进行*CVI*计算,前者是建立在各条目的“有效性、适用性、可行性或重要性”的基础上的,而后者则基于条目与研究内容之间的相关性,表示各条目能否反映相应的概念和维度。因此,在应用时应注意两者间的区别,不宜混淆使用。

3.3 正确计算并呈现*CVI*结果 *CVI*的计算以及根据计算结果对测量工具内容效度进行判断是内容效度评价的重要步骤,准确的计算和判断是内容效度检验可靠的表现。因此,详细、正确地呈现*CVI*计算和评价结果有助于读者获取有效信息。

本研究发现,在*CVI*的计算方面,仅少量研究先计算评定者间一致性(Inter-rater Agreement)^[6],在计算结果较好的情况下再对*CVI*进行计算;部分研究计算*CVI*的方法不正确,如有研究误将专家对某条目的评分相加后除以条目的最高可能总分作为条目水平内容效度指数(Item-level *CVI*, *I-CVI*)。在计算和计算结果呈现方面,少量未采用4分制评价的研究者未阐明计算*I-CVI*时选择评分几分以上的专家数除以专家总数;部分研究仅报道测量工具水平的内容效度指数(Scale-level *CVI*, *S-CVI*)结果,而未报道*I-CVI*的计算结果;较多研究未在摘要或正文中标识*S-CVI*的类型,仅简单表述为“量表总*CVI*”等;部分研究在评价测量工具时,其专家数量为5人或5人以下且*I-CVI*不为1.00,却描述该项测量工具的内容效度良好。

有学者建议,在计算*CVI*之前应先算出评定者间一致性,具体方法为被所有专家评分为1或2的条目数加上被所有专家评分为3或4的条目数除以条目总数^[7]。若评定者间一致性低于0.7,研究者应与专家沟通以确认专家间对函询表中测量概念和维度的定义是否一致,评分等级的设计是否恰当等;若评

定者间一致性不低于0.7,可进一步计算*CVI*^[2,6]。

*CVI*分为*I-CVI*和*S-CVI*。采用4分制时,*I-CVI*的计算方法为以单个条目为单位,评分为3或4的专家数除以专家总数^[1]。若研究者采用3分制或5分制,建议说明具体的计算方法。Lynn^[8]认为*I-CVI*的评价标准应根据专家人数而定:当专家人数 ≤ 5 人时,所有专家均认为该条目与相应的概念内容有较好的相关性,即*I-CVI*应为1.00,才认为这个条目的内容效度良好;而当专家人数 ≥ 6 人时,要求*I-CVI*的最小值为0.78^[2,4]。由于各个条目的*I-CVI*可能不同,所以建议研究者呈现每个条目的*I-CVI*情况或各条目的*I-CVI*数值范围。*S-CVI*有2种类型^[3],①全体一致*S-CVI*(*S-CVI/Universal Agreement*, *S-CVI/UA*)。*S-CVI/UA*是指被所有专家评为3或4分的条目数占全部条目数的百分比,反映的是专家全体一致认为相关的情况。若*S-CVI/UA*不低于0.8,则提示测量工具内容效度较好^[4]。*S-CVI/UA*的数值可随专家数量而变化,当专家人数增多时,意见不完全一致的可能性增大,因此,可能出现每个条目的*I-CVI*都较好,但*S-CVI/UA*较低的情况^[10]。②平均*S-CVI*(*S-CVI/average*, *S-CVI/Ave*)。*S-CVI/Ave*的计算方法有3种:测量工具所有*I-CVI*的均数;每个专家评分为3或4的条目占所有条目的比例的均数;评分为3或4出现的次数除以总评定数。这3种方法的结果均一致,多推荐使用第1种计算方法,因为其能对*I-CVI*做较好的承接,也便于理解^[2]。若*S-CVI/Ave*不低于0.90,则提示测量工具内容效度较好^[4]。由于对同一资料,2种*S-CVI*的计算结果可能相差较大,且相应的评判标准不同,所以建议研究者标明*S-CVI*类型或计算方式和所采用的评判标准。

此外,由于两个或多个专家对条目与相应维度的相关性评价一致可能是由于他们对选项的随机选择造成的^[11],因此有学者建议,应当对这种随机一致性(Chance Agreement)进行校正,计算调整后的Kappa值^[4]。Kappa值根据该条目的随机一致性和*I-CVI*计算获得,通常*I-CVI*越大,Kappa值也越高。校正随机一致性后,*I-CVI*不小于0.78,提示该测量工具的内容效度较好。

4 小结

*CVI*在护理测量工具开发领域应用广泛,但研究者在论文书写时往往对*CVI*相关内容描述得较简略,且存在一定的应用误区,主要在专家选择与描述、专家函询与评分、*CVI*的计算与描述这三方面存在问题。护理研究者应加强相关知识学习,正确应用该方法,如选择合适的专家小组成员、采用恰当的*CVI*评价表、保证*CVI*的计算正确等,并在论文书写时增加关键信息的描述,以加强测量工具开发的科学性和严谨性。