

复杂干预的设计和评价

刘森 洪志恒 詹思延

【导读】 介绍复杂干预研究的设计及过程,探索其在医学实践中的应用。在经典的干预性研究中,干预措施较为单一且效果评价较为简单。而目前越来越多的研究涉及到多个干预措施的综合,诸如心血管病的社区干预、卒中单元治疗等,此时复杂干预研究设计就较为适用。但复杂干预的设计、实施难度较大,效果评价也较困难。因此,复杂干预研究在设计、实施、评价和报告时,应严格遵守一定的标准,确保干预措施效果和研究的可重复性。

【关键词】 复杂干预; 设计; 评价

Development and evaluation of complex intervention LIU Miao, HONG Zhi-heng, ZHAN Si-yan.
Department of Epidemiology and Biostatistics, Peking University Health Science Center, Beijing 100191, China

Corresponding author: ZHAN Si-yan, Email: siyan-zhan@bjmu.edu.cn

This work was supported by a grant from the National Science and Technology Support Projects for the "Eleventh Five-Year Plan" of China (No. 2007BAI24B03).

【Introduction】 Experimental studies with only one intervention are not always sufficient for health care issnes, while the method of complex intervention is more useful. But the complex intervention is more complicated and needs carefully design and evaluation. This article will introduce the framework for design and evaluation of complex intervention and the revised version made by Medical Research Council, with five examples showing its application.

【Key words】 Complex intervention; Development; Evaluation

实验性研究是目前评估医学措施效果最可靠的科学方法^[1]。经典的实验性研究中干预措施比较单一,效果评价较为简单,如评价服用某药物后降低血脂的作用。而复杂干预研究(如基于社区的行为干预、卫生政策的评估等)的实施、评价、报告较为困难,在以往的研究中经常存在干预流程解释不清楚、过程评价不完整、报告不规范等问题。因此需要一种适宜、科学的方法对复杂干预设计进行严格规范,使得复杂干预的设计、实施、评估等规范化,利于读者理解和研究成果的推广。2000年英国医学研究理事会(Medical Research Council, MRC)制定的《复杂干预设计及评价框架》(框架)^[2]中提出,评价复杂干预的原则在于能否使复杂干预趋于标准化,且利于其他人接受和复制。框架将复杂干预按照药物试验的方法分为五期,并详细规定了每期的内容和要求。

DOI: 10.3760/cma.j.issn.0254-6450.2010.12.020
基金项目:“十一五”国家科技支撑计划(2007BAI24B03)
作者单位:100191 北京大学公共卫生学院 教育部流行病学重点实验室
通信作者:詹思延, Email: siyan-zhan@bjmu.edu.cn

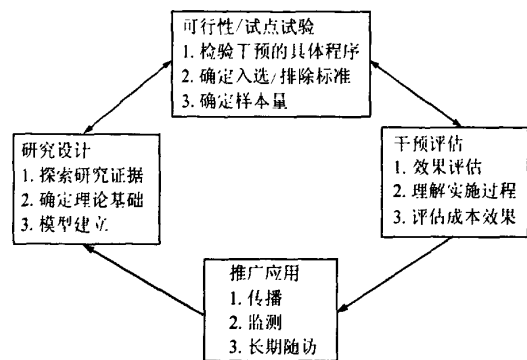
但一些研究者对其中部分内容提出疑问和建议,为此,2006年MRC成立专门的小组对框架进行更新,2008年出台新版本,针对复杂干预设计的分期、关键问题、研究实例、应用领域等问题进行扩展^[3,4]。

1. 复杂干预的定义和特点:框架中提出,复杂干预是由多种干预成分综合在一起的干预,这些成分彼此独立或相互影响。2008年在更新的框架中提出了复杂干预定义的多维性,以及在设计和评估过程中的复杂性。本研究总结了单一干预和复杂干预的异同(表1)。

表1 MRC 2008年框架中单一干预和复杂干预的异同

项目	单一干预	复杂干预
目标人群	确定	多样,根据实际情况可变
样本量	较少	相对较多
对照	安慰剂对照	非安慰剂对照,对照组用标准措施
干预措施	单一,确定	复杂,可变,干预成分间存在相互影响
分配隐匿	容易实施	较难实施
实施过程	简单	复杂,根据实际情况可变
是否遵循设计方案	遵循设计方案	不一定遵循方案,可适应实际情况
结局	单一,确定	多个结局,可变,复杂
因果链	直接,因果关系较强	复杂,因果关系多、较弱或不容易分辨
评价	较容易	较为复杂
应用	药物试验等	实际应用,如社区干预、政策评估

2. 复杂干预设计的过程及应用实例:2000 年的框架中采用药物试验分期方式将复杂干预的过程分为五期(理论假说、模型构建、探索性试验、验证实施和推广应用),并广泛应用于复杂干预的评价中^[5,6]。但也有研究者提出质疑,并不是所有的研究都是完全按照这个线性顺序进行,不同阶段间可能有重合或缺失。因此,在 2008 年更新的框架中,MRC 替换了原来的表示方法,将复杂干预设计分为 5 个阶段(设计、试点试验、评估、报告和推广应用),如图 1 所示。MRC 专门指出,尽管仍采用分阶段的方式表示复杂干预的过程,但在实际应用时不一定严格符合线性顺序。另外,报告存在于每一阶段,图中不再单独列出。



注:箭头表示不同时间期间存在相互影响

图 1 复杂干预设计和评估过程的关键成分(Craig 等^[3])

(1) 研究设计:研究者主要探索干预措施所能达到预期效果的理论基础和依据。首先需要确定研究的证据,如通过对以往研究进行系统综述,发现干预措施达到预期效果的证据;其次,要探索研究相关的理论,如通过查阅文献、与利益相关者访谈等,了解干预措施的原理、怎样达到预期效果等;最后,研究者要建立干预的模型,即了解干预的构成及各组分的相互关系。

Hardeman 等^[7]、Williams 等^[8]将心理学的理论和糖尿病流行情况结合起来,应用于糖尿病高危人群体力活动干预的随机试验研究。研究者首先通过文献回顾研究,总结了危险因素和结局的流行情况,确定目标人群、研究变量指标和结局指标。通过系统综述、与专家交流、专题小组访谈以及横断面调查等,获得了关于心理学理论应用于行为干预的理论基础,建立了干预效果评价和理论机制研究的模型。尽管研究最终的结局没有显著的效果,但在研究设计初期建立的理论模型,使得研究者能够非常清楚地评估干预措施的效果。

(2) 可行性/试点试验:通过可行性试点研究需要确定以下内容:①可接受性和可行性:利用收集的各种信息确定最适干预方案以及研究设计,包括不同形式干预的可行性和研究对象的可接受性,例如干预的强度、时间长短、研究对象的依从性等。②对照组采取的措施:在实验开始前应确定对照组采取的措施,可以是其他疗法干预、标准疗法干预或安慰剂对照,而不给对照组采取任何干预措施有伦理学与公平性的问题。一个解决办法是采用阶梯设计^[9],将研究对象干预的顺序进行随机化,并最终保证所有的研究对象都能接受干预措施。③研究设计的主要参数:试点试验也应尽量随机化,以评估效应的大小,最好能够得到研究设计的主要参数,以便进行样本量的计算。④结局指标:研究者在确定结局指标时不仅要包括与疾病相关的指标,还应该包括卫生保健系统相关的指标,如经济学指标。但结局指标的增多会增大研究者的工作量并耗费较多的成本,且对于研究的可接受性也是一个挑战。因此,通过试点研究选择恰当的结局指标非常重要。

Rudolf 等^[10]采用社区服务干预肥胖儿童减少早期发展为动脉硬化的危险。在正式试验前,研究者在两个社区开展了为期一年的试点研究。试点研究是基于社区的运动诊所,研究对象小组活动及与父母的小组活动等。并对试点工作进行分析,包括过程评估,如研究对象参与率、专家支持度,并收集对此次干预的看法;针对研究对象及其父母和祖父母进行的定性研究;测量研究对象 BMI 的改变、心理状况、生活质量等。研究者发现这种基于社区服务的儿童肥胖干预参与率高、效果好,可用于以后的随机化试验中。

(3) 复杂干预的评估:①研究方法的选择:研究者必须根据研究目的与实际选择合适的方法,比如对于社区干预、手术等问题,随机化试验设计就不适合。2008 年新版框架中提出如果研究不适合采用经典的个体随机对照试验,还可以采用群组随机试验、阶梯设计试验、偏好设计试验、N-of-1 设计试验等。如果是评估卫生政策,还可以采用非随机化试验等。MRC 指出,在复杂干预设计中,盲法的实施非常困难,尤其是在临床治疗、健康教育项目中。这可能会导致干预组与对照组的脱落差异,出现偏倚^[2]。②实施过程的理解:过程评估对于复杂干预的评价非常有价值。通过过程评估,既可以判断干预失败是由于干预措施无效还是实施过程的原因,也可以深入研究干预措施起作用的原理、确定

因果关系及结局的影响因素。③成本效果的评估：对于决策者来说非常重要，尤其在干预措施推广应用阶段。因此在研究设计初期，研究者就应探索研究的成本、干预产生的潜在效益等经济学问题。

Freemantle 等^[11]开展了一项研究，通过社区药师对全科医生进行教育，以鼓励按照指南开具 4 类常用处方药物（阿司匹林、ACEI、NSAIDs、抗抑郁药）。结果表明经过干预，医生按照指南开出 4 类药物的现象转好，约上升 5%；但是不同药物的上升率不同，最好的是阿司匹林，上升了 7%；最低的是 NSAIDs，上升了 2%。研究者想进一步通过对实施过程的评价来分析差异的原因。通过对药师和全科医生的调查以及小组讨论，研究者发现不同药物按照指南开具处方的差异与现实背景有关。如全科医生缺少超声心动检查等严格的证据以确诊缺血性心脏病患者，这与开具 ACEI 的处方有很大关系。而对于抗抑郁药，地方政府一般有管制，医师不能完全按照指南开处方。因此，在 Freemantle 等的研究中，4 类药物改变效果的不同不仅仅是因为干预措施的原因，也与实际环境有关。研究者通过仔细分析实施过程，提供了本次干预效果大小和变异的原因，这对于干预措施实际效果的评估以及读者的理解非常有帮助。

（4）推广应用：此阶段除了发表研究结果，更重要的是检验干预在实际环境中的应用，如干预在实际背景下的效果、干预的副反应等。以往的研究从试验到应用推广阶段存在很多困难，2008 年新版框架中提出了几条建议来促进实施，包括在研究选择与设计过程中咨询利益相关者的意见、探索各种背景下的应用证据、考虑研究的经济学指标、将研究结果最大化推广介绍、开展多方面的合作与交流等。

在试验研究中，由于时间和经费限制，不可能观察到所有的结局及可能出现的副反应。因此，当干预措施推广应用后，需要建立一个长期随访的系统，以监测干预实施的效果及可能的副反应。

Stead 等^[12]的研究说明了各种可能的因素都会影响干预措施的效果。以往试验研究显示尼古丁替代疗法（NRT）和安非他酮能够帮助吸烟者戒烟，且后者的长期效果更好。McEwen 和 West^[13]在两个戒烟机构实际应用了这两种戒烟方法并对效果进行了比较，通过 2626 名研究对象为期 2 年的效果观察，发现 NRT 的戒烟效果达到 42%，而安非他酮为 34%（ $P=0.003$ ）。那么是什么影响了两种干预措施的实际效果呢？研究者仔细分析了应用的过程，发现安

非他酮的实际操作过程复杂，且研究对象需要亲自到戒烟机构才能取到药物。而在 NRT 组中研究对象只需要提供一个凭单或信函即可收到药物，实施较为简单。正是由于安非他酮在实际操作中的繁复和附加条件，使其实际效果低于 NRT。

3. 讨论：复杂干预研究设计的优点是在一项研究中可以实施多个相互影响的干预措施，这更加符合研究者的实际需要。也正是这个优点使得该方法在设计、实施、评价、报告方面更加复杂。在实际应用中，复杂干预还存在以下方法学问题。

（1）学习曲线效应（learning-curve-effect），即干预效果随时间而增加^[2,4]。复杂干预必须考虑试验中是否有学习曲线效应的存在，如研究人员的技能在试验过程中是否不断提高？如果学习曲线效应存在，在正式招募试验者之前，须设定一个磨合阶段。国内有学者认为中医药临床疗效的评价中存在明显的学习曲线效应，不同的中医大夫对于中医辨证施治的领悟不同，需要反复的培训与实践才能保证中医大夫在治疗时的一致性^[5,6]。

（2）复杂干预过程描述不清楚，不易被读者理解。复杂干预研究设计的文章在讲述其具体干预实施方案时，由于干预措施较为复杂，即使花费了大量的文字，干预组与对照组的区别也未必能描述清楚。为此有学者在 2007 年提出使用流程图表示复杂干预的具体干预过程^[14]。图 2 用流程图表示新发心绞痛患者自我管理的干预实施过程（各行从上到下表示各个时间点干预措施的变化，不同列分别表示干预组和对照组），以图形式将研究设计结构清楚展现出来，易于读者理解^[15]。

时间轴	心绞痛计划干预组	健康教育对照组
随机化		
基线	a b c d	f g
1 周	e	
4 周	e	
8 周	e	
12 周	e	
6 月	结局测量	

注：不同字母代表不同的实施内容：a 患者填写调查表，看他们对咽痛的了解是否有误；b 护士与患者进行一次 30～40 min 的访谈；c 访谈时发给患者“咽痛计划”工作手册；d 发给患者录音磁带；e 护士给患者打电话（10～15 min）；f 宣传教育会上护士对患者讲解疾病的危险因素等；g 发给患者的书面材料

图 2 新发心绞痛患者自我管理干预实施过程（Perera 等^[14]）

（3）复杂干预设计的结局一般不止一个，因此数据分析较为困难。通常需要根据研究设计以及研究

目的来确定数据分析方法。MRC 提出,复杂设计中,针对维度较多的结局可以采用主成分分析来减少维度;有时间变量的结局可以采用时间序列分析或生存分析,多水平的资料可以采用多水平分析方法或混合效应模型;隐变量分析或结构方程模型可以帮助识别和建立干预与结局间的因果链条^[3,4]。

复杂干预设计的实施评价比较困难,数据分析较复杂,多种因素如盲法的实施、学习曲线效应等,对设计实施及效果评价有重要的影响。因此,研究者在应用复杂干预设计进行研究时,应参照 MRC 提出的框架进行设计、实施及评价。

参考文献

- [1] Blackwood B. Methodological issues in evaluating complex healthcare interventions. *J Adv Nurs*, 2006, 54(5): 612-622.
- [2] MRC. A framework for development and evaluation of RCTs for complex interventions to improve health, 2000.
- [3] Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new medical research council guidance. *BMJ*, 2008, 337: a1655.
- [4] Anderson R. New MRC guidance on evaluating complex interventions. *BMJ*, 2008, 337: a1937.
- [5] Zhi YJ, Xie YM, Weng WL, et al. Modeling of TCM clinical complex intervention evaluation. *World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica*, 2007, 9(6): 25-30. (in Chinese)
支英杰, 谢雁鸣, 翁维良, 等. 结合中医辨证论治的特点构建中医临床复杂干预评价模型. *世界科学技术-中医药现代化*, 2007, 9(6): 25-30.
- [6] Xie YM, Zhi YJ, Wang YY. New methods on evaluation of TCM clinical efficacy-complex intervention clinical efficacy evaluation. *J Tradit Chin Med*, 2008, 49(5): 395-397. (in Chinese)
谢雁鸣, 支英杰, 王永炎. 适合中医临床疗效评价的新法初探——复杂干预措施的临床疗效评价方法. *中医杂志*, 2008,

49(5): 395-397.

- [7] Hardeman W, Sutton S, Griffin S, et al. A causal modelling approach to the development of theory-based behaviour change programmes for trial evaluation. *Health Educ Res*, 2005, 20(6): 676-687.
- [8] Williams K, Prevost AT, Griffin S, et al. The proactive trial protocol — a randomised controlled trial of the efficacy of a family-based, domiciliary intervention programme to increase physical activity among individuals at high risk of diabetes [ISRCTN61323766]. *BMC Public Health*, 2004, 4: 48.
- [9] Yang ZY, Zhan SY. Application of “stepped-wedge design” methodology in randomized controlled trials. *Chin J Epidemiol*, 2010, 31(1): 92-95. (in Chinese)
杨祖耀, 詹思延. 阶梯设计在随机对照试验中的应用. *中华流行病学杂志*, 2010, 31(1): 92-95.
- [10] Rudolf M, Christie D, McElhone S, et al. WATCH IT: a community based programme for obese children and adolescents. *Arch Dis Child*, 2006, 91(9): 736-739.
- [11] Freemantle N, Nazareth I, Eccles M, et al. A randomised controlled trial of the effect of educational outreach by community pharmacists on prescribing in UK general practice. *Br J Gen Pract*, 2002, 52(477): 290-295.
- [12] Stead LF, Perera R, Bullen C, et al. Nicotine replacement therapy for smoking cessation. *Cochrane Database Syst Rev*, 2008, (1): D146.
- [13] McEwen A, West R. Do implementation issues influence the effectiveness of medications? The case of nicotine replacement therapy and bupropion in UK Stop Smoking Services. *BMC Public Health*, 2009, 9: 28.
- [14] Perera R, Heneghan C, Yuditkin P. Graphical method for depicting randomised trials of complex interventions. *BMJ*, 2007, 334(7585): 127-129.
- [15] Lewin RJ, Furze G, Robinson J, et al. A randomised controlled trial of a self-management plan for patients with newly diagnosed angina. *Br J Gen Pract*, 2002, 52(476): 194-196, 199-201.

(收稿日期: 2010-04-13)

(本文编辑: 张林东)

• 书讯 •

《暴力流行病学》现已出版

王声湧和林汉生教授主编的《暴力流行病学》已由人民卫生出版社出版发行。这部以暴力为主题的教科书,在医药卫生部门还是首次问世,也是全球第一部论述暴力的流行病学。该著作是继《伤害流行病学》和《伤害流行病学现场研究方法》之后,由人民卫生出版社实现王声湧教授“伤害和暴力控制三部曲”的宏愿,为我国伤害预防控制学科的发展起到导向性作用。

《暴力流行病学》详细介绍常见的暴力行为,大体分为人际间暴力、自我暴力和集团暴力三部分。人际间暴力围绕儿童虐待和校园暴力、虐待老人、妇女暴力和性暴力、医院工作场所暴力四个方面论述;自我暴力针对自杀的流行特征、自杀的危险因素、自杀的预防与控制及自杀危机的干预进行讨论;在集团暴力中介绍了集团暴力的定义和类型、近代常见的集团暴力及其应急管理 with 紧急救援,同时附加了近年来各地常发生的群体性事件;在暴力的研究方法一章中,著者详细介绍了运用流行病学和统计学方法在暴力研究上的心得和体会。

《暴力流行病学》中有相关的理论辨析、国内外现状描述、预防干预措施、案例分析和研究方法介绍。内容深入浅出,医学、社会学、心理学和哲学知识相映成辉、水乳交融,可供预防医学人员、医务工作者、卫生工作管理者和医学院校师生作为教科书、参考书,同时也是大众读者值得品味的常识读物。均可在人民卫生出版社网站和各购书网站以及各地书店订购。每本定价 36 元。

(王声湧 供稿)