

· 综 述 ·

阶梯设计试验研究方法及其报告规范*



司华新, 乔晓霞, 金雅茹, 王翠丽

【摘 要】阶梯设计是在服务提供和政策评估领域出现的一种新兴的整群随机对照试验设计, 近年来逐渐受到医疗卫生领域研究者的重视。阶梯设计试验不需要设置专门的对照组, 在资源有限只能分步实施干预的情况下仍保持传统随机对照试验的优势。本文从基本原理、适用情况、特点、设计类型、数据统计分析方法和相关报告规范等方面对阶梯设计试验进行了综述, 旨在为更好地开展阶梯设计试验提供参考并准确评估服务项目在真实世界的实际效果。

【关键词】阶梯设计试验; 研究方法; 报告规范

中图分类号:R 195 文献标识码:A 文章编号:1001-0580(2022)04-0503-07 DOI:10.11847/zggws1134273

Method and reporting guideline for stepped-wedge trials

SI Hua-xin, QIAO Xiao-xia, JIN Ya-ru, et al (School of Nursing, Peking University, Beijing 100191, China)

【Abstract】 Stepped-wedge trial is an emerging cluster randomized controlled trial in the field of service provision and policy evaluation, which is becoming increasingly popular among health care researchers in recent years. There is no need to set up a special control group for stepped-wedge trials, and the advantage of traditional randomized controlled trial is still maintained in the case that the intervention can only be carried out step by step due to limited research resources. In the study, we reviewed the rationale, applicable conditions, characteristics, design types, statistical analysis methods, and relevant reporting guidelines of stepped-wedge trials for providing a reference to quality improvement in stepped-wedge trials and accurate evaluation on health service in real practice.

【Key words】 stepped-wedge trial; research method; reporting guideline

世界卫生组织呼吁开展实施研究以应对如何在真实世界中实施经过验证的干预措施的挑战, 帮助将理论中的可能性转化为日常实践, 从而提高卫生服务质量^[1-2]。而阶梯设计试验作为一种真正意义上的实施研究设计^[3], 近年来逐渐受到医疗卫生领域研究者的重视。阶梯设计是一种新兴的实用型整群随机对照试验设计, 各群组随机在不同时间点实施干预, 最终所有组均将接受干预, 不需要设置专门的对照组, 在资源有限只能分步实施干预的情况下仍保持传统随机对照试验的优势^[4-7]。20 世纪 80 年代开展的冈比亚肝炎干预研究是最早也是最广为人知的阶梯设计试验研究^[8], 随后尤其是 2007 年 Hussey 等^[5]发表了关于阶梯设计试验样本量计算及分析方法的文章以来, 阶梯设计的应用逐渐增多, 主要集中于医疗卫生服务和公共卫生政策评价等领域^[9-11]。然而, 一篇系统综述指出目前阶梯设计试验研究多集中于高收入国家, 如英国、澳大利亚、荷兰等^[12], 国内关于阶梯设计试验研究则较少, 仅 2 篇实证研究分别在急性冠脉综合征患者^[13]和自闭症患儿^[14]中开展了阶梯设计随机对照试验。另有 3 篇方法学研究虽对阶梯设计的基本原理、优缺点、设计类型以及样本量和功效的计算方

法进行了介绍^[15-17], 但缺乏对其数据统计分析方法及报告规范的介绍, 且以往研究介绍的阶梯设计类型(固定队列阶梯设计、开放队列阶梯设计、连续入组短期暴露阶梯设计)是根据研究对象类型划分的^[17], 并未从阶梯设计方法本身角度考虑。为了对阶梯设计试验更好地开展提供参考依据, 能更准确地评估服务项目在真实世界的实际效果, 本文从阶梯设计的基本原理、适用情况、特点、设计类型、数据统计分析方法和相关报告规范等方面进行了综述如下。

1 阶梯设计试验概述

1.1 阶梯设计试验基本原理 (图 1) 在阶梯设计试验中, 干预措施是在一定时间内按照顺序实施的, 即在每一个时间点均会有新的群组接受干预, 不同群组接受干预的顺序是随机分配的, 在随机分配结束时所有群组都将接受干预^[6]。在最常见的阶梯设计方案中, 所有群组在对照状态下开始, 在干预状态下结束, 并在均匀间隔的时段交叉, 这在一定程度上模仿了真实情况下(非实验性)干预措施的实施过程^[18]。阶梯设计试验的基本原理类似于交叉设计^[5], 即不同的群组在不同的时间点(阶段)

* 基金项目: 国家自然科学基金(71673168)
作者单位: 北京大学护理学院, 北京 100191
作者简介: 司华新(1994-), 女, 河北衡水人, 博士在读, 研究方向: 老年护理。
通信作者: 王翠丽, E-mail: cwangpk@163.com

交叉(切换处理),但通常只在一个方向上交叉,即从对照状态到干预状态。第一个时间点通常对应于一次基线测量,在此阶段无群组接受干预措施,在随后的时间点群组开始接受干预,并观测相应的结局指标(图 1)。在一个时间点上可能有一个或多个群组开始接受干预,但一个或多个群组何时开始干预的时间是随机的。

区块/ 群组	时间				
	0	1	2	3	4
1					
2					
3					
4					

注:白色框表示该群组尚未接受干预,处于对照状态;黑色框表示该群组接受干预,处于干预状态。

图 1 阶梯设计试验基本设计示意图

1.2 阶梯设计试验适用情况 阶梯设计试验的适用情况为:(1)希望采用随机对照试验方法评估某一干预措施的效果,但由于人力、物力、财力等资源限制只能分阶段开展时;(2)希望边推行边改进某项干预措施时(阶梯设计允许研究者在前一阶段的干预实施过程中学习经验);(3)当已知某种干预措施有效且利大于弊,但不对研究参与者进行干预不符合伦理学道德时;(4)当群组数量较少,但群内相关系数较高且群组规模较大时^[7]。

1.3 阶梯设计试验的特点(表 1) 与整群随机对照试验类似^[19],阶梯设计试验研究的单位多为医院、社区、诊所、学校、家庭和工作场所等,因而该试验亦包括群体和个体 2 个水平,这也是它们与传统随机对照试验的最大区别。为此,阶梯设计试验在设计、研究对象纳入、样本量计算和统计分析等

方面既要考虑个体水平,又要考虑群体水平,且还需要考虑阶段数、群组间的相关性和时间效应等,具体见表 1。

1.4 阶梯设计试验的优点与局限性 阶梯设计试验的优点包括:(1)允许干预措施分阶段实施,节省人力、物力,能解决现场实施的可行性问题;(2)在阶梯设计研究中,所有研究参与者均接受干预措施,保证公平性,避免了伦理学问题;(3)阶梯设计试验无需设立专门的对照组,而是将群组作为自己的对照,在保持随机对照试验优势的同时亦避免了传统随机对照试验中被随机分配到对照组的参与者因“失望效应”而引发的不依从或退出,提高了统计效能^[4,12];(4)阶梯设计可以发现和控制时间效应,即阶梯设计研究的目的不仅可以评估干预效果,而且还可以确定干预时间(在早期干预与晚期干预的极端情况下)是否会影响干预效果^[5],这种时间效应包括季节、卫生条件等环境变化和疾病自身的发生发展变化等情况,随着试验的推进,此种时间效应可能会给干预效果的评价造成偏倚^[21-22],而采用阶梯设计则有助于明确是否存在此种时间效应,并对其进行调整;(5)当群组间组内相关系数较大时,阶梯设计试验较平行整群随机对照试验的统计效能更高^[4,23]。但是,阶梯设计试验在设计、组织实施和统计分析过程中仍存在一定的局限性:(1)与传统的平行整群随机对照试验相比,阶梯设计试验需要设计多个时间点,且考虑到治疗效应延迟需要将每一个时间段的长度定得足够长,因而其试验周期可能更长且存在一定的测量负担^[5-6];(2)在阶梯设计实施过程中,可能会出现群组内沾染,如干预措施的嵌入较计划时间更长,或对长期暴露于干预状态下参与者的结局测评延迟,会影响结果的真实性^[7];(3)在干预实施过程中,试验参与者和干预提供者均会意识到从无干预到有干预状态的改变,因而不可能对其实施盲法,但可通过对结局评估者实施盲法来减少信息偏倚;(4)与整群随机对照试验类似,阶梯设计试验的数据分析需综合考虑群体水平和个体水平,且需考虑时间效应、群组内相关性和参与者重复测量的相关性,统计方法更为复杂^[21,24]。

表 1 阶梯设计试验与传统随机对照试验、整群随机设计试验的区别

试验类型	研究设计	样本量计算	统计分析
传统随机对照试验	以个体为基础,随机化方案、干预措施等均基于个体	考虑效应、Ⅰ类错误、Ⅱ类错误	不需要考虑群集性,分析方法较为简单
整群随机对照试验	以群组为基础,随机化方案、干预措施等均基于群组	还需考虑群组数、群组大小、群内相关系数,样本量往往更大	考虑群集性,分析方法更加复杂;比较个体和群组水平的基线资料;需给出结局的群内相关系数
阶梯设计试验	多以群组为基础,随机化方案、干预措施等均基于群组	另外需考虑阶段数、群组间的相关性、时间和延迟效应,样本量较整群随机对照试验小 ^[20]	可采用同期比较和前后比较2种方法;在前后比较中,还需考虑结局指标的时间变化趋势;群组内的相关性和多次重复测量间的相关性也需考虑

2 阶梯设计试验类型

根据阶梯设计试验实施过程中不同时间点群组的观测数据收集或使用与否, Hemming 等^[4, 25]将阶梯设计分为完整阶梯设计和不完整阶梯设计。

2.1 完整阶梯设计 (图 2) 完整阶梯设计指在每个固定的时间点上均有固定数量的群组(通过随机选择)开始接受干预, 传统的阶梯设计试验就是一个完整阶梯设计, 即在一定数量时间点上的一组群组连续随机地接受干预, 且在随后的所有时间点均保持暴露于干预状态, 并在基线时点和每个时间点均进行观测, 形成用于分析的数据。图 2 为完整阶梯设计试验示例, 图中共有 6 个时间点(包括 1 个基线时点), 在 5 个(1~5)时间点上分别有不同的群组被随机分配接受干预, 第 1 个时间点(0)用于基线数据收集。方框中数字为“1”表示在该时点该区块暴露于干预措施, 方框中数字为“0”表示在该时点该区块不接受干预措施(对照状态)。这些区块可以由 ≥ 1 个群组构成, 传统的完整阶梯设计假设每个区块中包含相同数量的群组^[26]。在完整阶梯设计试验中, 所有区块内的群组在所有时点上均会收集数据, 并未考虑试验实施过程中可能存在的数据收集限制, 也可能会因为多次随访测量对研究参与者造成测量负担^[27]; 此外, 其假定所有群组不是处于干预状态就是处于对照状态, 而忽略了从对照状态到干预状态之间的过渡期, 可能会过高估计统计效能。

区块/ 群组	时间					
	0	1	2	3	4	5
1	0	1	1	1	1	1
2	0	0	1	1	1	1
3	0	0	0	1	1	1
4	0	0	0	0	1	1
5	0	0	0	0	0	1

注: “0”表示对照状态; “1”表示干预状态。

图 2 完整阶梯设计试验示例

2.2 不完整阶梯设计试验 (图 3、4) 考虑到实际应用过程中可能会出现阶梯设计试验实施步骤不完整的情况, Hemming 等^[4]对完整阶梯设计做出一些试验设计上的变化, 称为不完整阶梯设计, 即在某些步骤(实施或过渡阶段)中对某些群组不进

行数据收集。这些特征是有可能在临床实践中发生的、较为实用的设计变化, 且近年来有越来越多的研究开始采用不完整阶梯设计方法^[27-32]。不完整阶梯设计是完整阶梯设计的变形, 这些不完整的设计通常只包括由实施阶段带来的不完整, 如, 在某些时间点无群组开始接受干预措施, 在某些时间点(过渡期)未收集某个群组的观测数据, 在不同时间点上随机分配给干预组的群组数量不同等。图 3 展示的是一个不完整阶梯设计, 在该设计中某个特定的时间点, 群组可能是暴露于干预状态(1), 也可能是处于非干预状态(0), 或者不收集观测数据(·)。在此不完整阶梯设计实施过程中, 群组随机接受干预, 但数据观测与收集仅在干预实施前的 1 个时间点和干预实施后的 2 个时间点进行。此类型的不完整阶梯设计减少了数据收集总量, 尤其是减少了同时的数据收集量, 考虑了由于数据收集限制而导致的试验实施中的实际限制, 也可以在一定程度上避免测量负担。然而, 在群组数量固定的情况下, 此种数据收集方法会降低统计效能, 而通过增加干预实施前后的数据收集量、样本量及群组数量可以减弱其对统计效能的影响^[26, 33]。图 4 也是一种不完整阶梯设计, 即在试验实施的整个观察期均收集观测数据, 但在开始实施干预的时间点不进行数据收集, 称为伴有过渡期的不完整阶梯设计。此种不完整阶梯设计进一步考虑了干预实施阶段研究对象从未干预到干预期的铺垫状态, 可以避免由干预措施嵌入时间较原计划长而导致的群组内沾染^[7]; 且在此过渡期内, 研究者可以考虑试验场所(社区、医院等)或群组在干预实施过程中可能需要的具体优化策略以及时间安排等, 以保证后期干预措施的顺利实施^[26]。

区块/ 群组	时间					
	0	1	2	3	4	5
1	0	1	1	·	·	·
2	·	0	1	1	·	·
3	·	·	0	1	1	·
4	·	·	·	0	1	1

注: “0”表示对照状态; “1”表示干预状态; “·”表示不收集观测数据。

图 3 不完整阶梯设计试验示例(干预前后分别进行 1 次和 2 次数据收集)

区块/ 群组	时间					
	0	1	2	3	4	5
1	0	·	1	1	1	1
2	0	0	·	1	1	1
3	0	0	0	·	1	1
4	0	0	0	0	·	1

注：“0”表示对照状态；“1”表示干预状态；“·”表示该时点不收集观测数据。

图 4 伴有过渡期的不完整阶梯设计试验示例

3 阶梯设计试验统计分析方法（表 2）

阶梯设计试验的统计分析方法亦存在争议,关于阶梯设计试验的系统综述表明目前研究中使用的数据统计方法有较大差异,从简单的横断面统计方法(如 *t* 检验和 Manne-Whitney *U* 检验)到更为复杂的方法(如混合模型)^[9, 34]。且对阶梯设计试验数据进行分析时需要考虑的最重要的特征是单向交叉设计,即干预效果一部分可以通过被试者自身前后对照(被试者在某个时间点由对照转变为干预状态)进行测量,另一部分可以通过被试者间的比较(在某个时间点时干预组与对照组进行比较)进行测量^[21]。在理想情况下,此两部分的干预效果应结合在一起进行分析,为此对阶梯设计试验的数据进行分析时有必要采用一种能够综合这些效果的方法,即混合模型^[24, 35]。与传统的随机对照试验不同,在阶梯设计试验数据分析过程中,时间变量也发挥着重要作用,会影响干预效果的估计^[22]。此外,还应考虑对结局变量的基线差异进行调整^[36]。基于以上考虑, Twisk 等^[21] 总结出 4 种阶梯设计试验统计分析方法。方法 1 是将所有干预时点与对照时点的测量数据进行比较。在此方法中,干预变量是 1 个与时间相关的二分类变量,估计的干预效果反映所有干预时点测量数据与所有对照时点测量数据间的差异。由于干预效果仅由 1 个数字结果反映,此种方法不能回答长时间干预与短时间干预是否存在差异的问题。在方法 2 中,干预变量是 1 个与时间无关的分类变量,通过比较不同组别之间的差异来评估干预效果,每个组别均为干预时点和对照时点测量的不同组合。在方法 3 中,干预变量是 1 个与时间相关的分类变量,将接受不同干预时长群组的测量数据分别与所有对照时点的测量数据进行比较。方法 3 是方法 1 的延伸,将方法 1 中的

干预组按照干预时长进行了分组。方法 2 和方法 3 可以检验干预时长对干预效果的影响,方法 2 的分析结果可以解释为接受不同干预时长组别的平均差异,但这些差异反映的只是被试间的干预效果;方法 3 的分析结果可以更为直接地显示不同干预时长对干预效果的影响,同时反映被试间和被试内两部分的干预效果,但此方法的缺点是随着干预周期的延长,接受越长时间干预措施的群组数或被试人数就越少,可能会降低其统计效能^[37-38]。方法 4 与前 3 种方法略有不同,它不是分析结局变量在不同时点的观测值,而是分析结局变量在连续 2 个测量时点前后间(转换期)的变化,然后比较被试者从对照期转换为对照期、被试者从对照期转换为干预期以及被试者从干预期转换为干预期 3 种转换期之间的差异。由于此方法以连续测量前后之间的变化为结局变量,因而被试者内部的相关性几乎为零^[22],仅评估了被试内的干预效果。以上 4 种统计分析方法分别揭示了阶梯设计试验干预效果的不同方面,统计分析方法的选择在一定程度上取决于干预类型及干预效果与时间相关的可能性。表 2 为对 4 种阶梯设计统计分析方法优缺点的总结。当干预效果与时间无关时,方法 3 可能是对干预效果的最佳估计,因为此方法结合了被试间效应和被试内效应,且考虑了干预时长。然而,当干预效果与干预起始时间相关时,方法 3 不能对干预效果进行有效地估计,最稳定的结果可以由方法 2 得出,这是因为方法 2 中分析结果不受时间影响(属于某个特定组别与时间无关,因而时间不是混杂因素),但缺点是此方法仅估计了被试间的干预效果。此外, Twisk 等^[21] 建议将不同分析方法结合起来以获得有意义、可解释的结果。

表 2 四种阶梯设计统计分析方法的优缺点

优缺点	方法1	方法2	方法3	方法4
是否考虑被试间和/或被试内效应?	两者均有	仅被试内效应	两者均有	多为被试内效应
是否需要调整时间变量?	是	否	是	是
是否能分析干预时长的影响?	否	是	是	否
是否需要调整基线时的结局测量?	是	是	是	是
是否需要假设干预效果与干预开始时间无关?	是	否	是	是
能否评估干预效果的延迟性?	否	是	是	是(部分)

4 阶梯设计试验报告规范（表 3）

阶梯设计本身包含的一些特性,如潜在的时间混杂因素、群组相关性随时间的变化、群组内沾染的可能性、干预效果随时间变化的可能性、多种设计类型的变异等^[39-42],增加了阶梯设计试验报告

的复杂程度^[24]。目前,关于阶梯设计试验报告充分性的系统综述发现以往研究在报告阶梯设计试验时存在许多不足之处,包括缺乏设计的基本细节、使用术语不一致、调整时间效应的相关报告缺乏清晰度以及经常不报告伦理审查和试验注册等^[10-11,43-44],提示有必要为阶梯设计试验制定一个具体的报告指南,由此 Hemming 等^[45-46]将为报告个体随机对

照试验而开发的 CONSORT 声明扩展应用于阶梯设计试验中。扩展的 CONSORT 声明考虑了阶梯设计试验的特殊性,共包含 26 个条目(见表 3),有些条目是 CONSORT 声明中原有的,有些是经过修改的,还有一些是新添加的,Hemming 等^[46]希望通过制定此报告规范以减少未来阶梯设计试验研究报告质量低下的情况。

表 3 阶梯设计整群随机试验 CONSORT 扩展声明条目清单

主题		条目	描述	
标题与摘要		1a	标题能识别为阶梯设计整群随机试验	
		1b	结构化摘要,包括试验设计、方法、结果和结论	
前言	背景与目标	2a	科学背景,使用整群设计和阶梯设计的原理	
		2b	具体目的或假设	
方法	试验设计	3a	利用文字和图表描述试验设计,包括群组的定义、序列数、随机分配到每个序列的群组数量、观察周期数、每一周期之间的时间间隔以及不同时期的参与者是相同的人/不同的人/两者的混合	
		3b	试验开始后对试验方法所做出的重要改变(如入选标准),并阐明理由	
	研究对象	4a	研究对象和群组的纳入标准	
		4b	数据收集的场所和地点	
	干预措施	5	详细描述干预措施和对照措施,包括干预措施是维持还是重复以及干预措施是在群体水平、个体水平还是两者均提供,以使他人能够重复	
	结局指标	6a	明确定义主要和次要结局指标,并说明它们在何时、如何进行测评	
		6b	试验开始后观测的结局指标是否有改变,并阐明理由	
	样本大小	7a	样本量大小如何确定,详细描述计算方法及相关参数(显著性水平、统计效能、群组大小、群组数量等),便于重复计算,并对不同时期的群内相关性作出假设	
		7b	必要时,描述进行中期分析的目的、时间、方法及试验中止的条件	
	随机方法	序列的产生	8a	产生随机分配序列的方法
8b			随机方法的类型;详细阐明限定随机或分层随机方法(如果使用)	
分配隐藏		9	详细说明分配是基于群组而不是个体以及招募结束前在群组中进行的任何分配隐藏方法	
实施		10a	谁产生随机分配序列,谁招募群组,谁分配群组到相应的序列	
		10b	描述群组中参与者的纳入方法(如完全枚举、随机抽样、连续招募、在一个固定的时间点招募),包括谁招募或认定参与者	
		10c	是否、何时、从谁那里、因为什么征得知情同意;干预和对照状态下知情同意内容是否有区别	
盲法		11a	如果实施了盲法,在分配序列后谁被实施了盲法(如群体水平的参与者、个体水平的参与者、结局指标测评者),且盲法是如何实施的	
		11b	如有必要,描述干预措施与对照措施的相似性	
统计方法		12a	用于比较各组主要和次要结局指标的统计方法,包括如何考虑时间效应、群组和重复测量	
		12b	附加分析的方法,如亚组分析、敏感性分析和校正分析	
结果	研究对象的流程 (使用流程图)	13a	报告每组或各分配的序列上符合纳入标准、被随机分配、接受干预并用于分析主要结局的群组和参与者的数量	
		13b	报告每组或各分配的序列上群组和参与者的失访和排除情况及原因	
	研究对象的招募	14a	报告招募研究对象和随访时间的长短,说明具体日期;报告每一阶段的日期、干预实施起始日期及偏离原定计划日期的情况	
		14b	阐明试验结束或中止的原因	
	基线数据	15	每组或各分配的序列上个体水平和群组水平的基线特征	
	分析的数量	16	纳入每一种分析的群组和研究对象的数量以及分析是否按照原计划进行	
	结局和估计	17a	对每个主要和次要结局指标报告各组的结果、效应估计值及精确度(如95 % <i>CI</i>),报告分析中估计的相关系数(或协方差)和时间效应	
		17b	对于二分类结局变量,建议同时提供绝对效应值和相对效应值	
	辅助分析	18	报告进行的其他所有分析,包括亚组分析和校正分析,阐明哪些是预先设定的分析,哪些是临时添加的分析	
	不良反应事件	19	各组出现的严重不良事件或副作用	
	讨论	局限性	20	试验的局限性,报告潜在偏倚和不精确的原因以及出现多种分析结果的原因(如有必要)
		可推广性	21	试验结果向个体和(或)群组的可推广性(外部效度、适用性)
	其他信息	解释	22	与结果相对应的解释,权衡试验结果的利弊,并考虑其他相关证据
		注册	23	试验注册号及注册机构名称
试验方案		24	如果可以的话,在哪里可以获取完整的试验方案	
资助		25	资助和其他支持(如提供药品)的来源以及资助者的作用	
伦理		26	是否得到伦理审查委员会的批准,报告伦理审查和知情同意的细节	

5 小 结

阶梯设计是在服务提供和政策评估领域出现的一种新兴的整群随机对照试验设计,该设计方法适用于评价“利大于弊”的干预措施,其在保持随机对照原则的基础上,可以避免由资源受限无法全面开展干预的问题。在临床实践中,对阶梯设计试验干预措施的检验或实施可以作为对常规服务的补充,从而最大限度地提高试验结果的可行性和推广性。研究者可采用阶梯设计方法在一定范围内全面推行某服务项目,同时通过随机对照试验方案评价该项目的应用效果,但阶梯设计试验在方案设计、实施、统计分析 & 报告方面均较为复杂,研究者需要结合资源、卫生保健设施等实际情况进行严谨的试验方案设计与实施,根据设计类型选择合适的数据统计分析方法,并进行规范报告,从而能够保证准确评估服务项目在现实世界的实际效果。

志谢 感谢北京大学公共卫生学院詹思延教授对本文撰写提出的宝贵建议

利益冲突 所有作者声明不存在利益冲突

出版授权 作者同意以纸质版和网络版的形式同时出版

参考文献

- [1] Peters DH, Adam T, Alonge O, et al. Republished research: implementation research: what it is and how to do it[J]. *British Journal of Sports Medicine*, 2014, 48(8): 731 – 736.
- [2] 刘万奇, 杨金侠, 汪志豪, 等. 中国基本公共卫生服务实施经验、问题与挑战 [J]. *中国公共卫生*, 2020, 36(12): 1677 – 1681.
- [3] 谢润生, 徐东, 李慧, 等. 医疗卫生领域中实施科学的研究方法 [J]. *中国循证医学杂志*, 2020, 20(9): 1104 – 1110.
- [4] Hemming K, Lilford R, Grling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs[J]. *Statistics in Medicine*, 2015, 34(2): 181 – 196.
- [5] Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials[J]. *Contemporary Clinical Trials*, 2007, 28(2): 182 – 191.
- [6] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review[J]. *BMC Medical Research Methodology*, 2006, 6: 54.
- [7] Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice?[J]. *International Journal of Epidemiology*, 2020, 49(3): 1043 – 1052.
- [8] The Gambia Hepatitis Study Group. The Gambia hepatitis intervention study[J]. *Cancer Research*, 1987, 47(21): 5782 – 5787.
- [9] Mdege ND, Man MS, Taylor CA, et al. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation[J]. *Journal of Clinical Epidemiology*, 2011, 64(9): 936 – 948.
- [10] Grayling MJ, Wason JMS, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features[J]. *Trials*, 2017, 18(1): 33.

- [11] Taljaard M, Hemming K, Shah L, et al. Inadequacy of ethical conduct and reporting of stepped wedge cluster randomized trials: results from a systematic review[J]. *Clinical Trials*, 2017, 14(4): 333 – 341.
- [12] Beard E, Lewis JJ, Copas A, et al. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014[J]. *Trials*, 2015, 16: 353.
- [13] Wu YF, Li SS, Patel A, et al. Effect of a quality of care improvement initiative in patients with acute coronary syndrome in resource-constrained hospitals in China: a randomized clinical trial[J]. *JAMA Cardiology*, 2019, 4(5): 418 – 427.
- [14] So WC, Cheng CH, Law WW, et al. Robot dramas may improve joint attention of Chinese speaking low-functioning children with autism: stepped wedge trials[J]. *Disability and Rehabilitation: Assistive Technology*, 2020. doi: 10.1080/17483107.2020.1841836.
- [15] 杨祖耀, 詹思延. 阶梯设计在随机对照试验中的应用 [J]. *中华流行病学杂志*, 2010, 31(1): 92 – 95.
- [16] 闫翔宇, 苏鹤轩, 张波, 等. 阶梯整群随机对照试验的样本量和功效计算方法 [J]. *现代预防医学*, 2019, 46(23): 4229 – 4232, 4246.
- [17] 王彬, 闫世艳, 马利. 阶梯设计方法及其在中医临床研究中的应用 [J]. *中医杂志*, 2016, 57(17): 1481 – 1485.
- [18] Hooper R, Eldridge SM. Cutting edge or blunt instrument: how to decide if a stepped wedge design is right for you[J]. *BMJ Quality and Safety*, 2020, 30(3): 245 – 250.
- [19] 陈新林, 莫传伟, 徐谦, 等. 整群随机试验的设计、统计分析方法及应用 [J]. *中国循证医学杂志*, 2015, 15(6): 741 – 744.
- [20] Woertman W, de Hoop E, Moerbeek M, et al. Stepped wedge designs could reduce the required sample size in cluster randomized trials[J]. *Journal of Clinical Epidemiology*, 2013, 66(7): 752 – 758.
- [21] Twisk JWR, Hoogendijk EO, Zwijsen SA, et al. Different methods to analyze stepped wedge trial designs revealed different aspects of intervention effects[J]. *Journal of Clinical Epidemiology*, 2016, 72: 75 – 83.
- [22] Twisk JWR. Applied longitudinal data analysis for epidemiology: a practical guide[M]. 2nd ed. Cambridge, UK: Cambridge University Press, 2013.
- [23] Grling AJ. Relative efficiency of unequal cluster sizes in stepped wedge and other trial designs under longitudinal or cross-sectional sampling[J]. *Statistics in Medicine*, 2018, 37(30): 4652 – 4664.
- [24] Hemming K, Haines TP, Chilton PJ, et al. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting[J]. *BMJ*, 2015, 350: h391.
- [25] Hemming K, Grling A. A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials[J]. *The Stata Journal*, 2014, 14(2): 363 – 380.
- [26] Unni RR, Lee SF, Thabane L, et al. Variations in stepped-wedge cluster randomized trial design: insights from the Patient-Centered Care Transitions in Heart Failure trial[J]. *American Heart Journal*, 2020, 220: 116 – 126.
- [27] Dupuis M, Kuczewski E, Villeneuve L, et al. Age Nutrition Chirurgie (ANC) study: impact of a geriatric intervention on the screening and management of undernutrition in elderly patients operated on for colon cancer, a stepped wedge controlled trial[J]. *BMC Geriatrics*, 2017, 17(1): 10.

[28] Treichler EBH, Glorioso D, Lee EE, et al. A pragmatic trial of a group intervention in senior housing communities to increase resilience[J]. [International Psychogeriatrics](#), 2020, 32(2): 173 – 182.

[29] Haines TP, Palmer AJ, Tierney P, et al. A new model of care and in-house general practitioners for residential aged care facilities: a stepped wedge, cluster randomised trial[J]. [Medical Journal of Australia](#), 2020, 212(9): 409 – 415.

[30] Walker P, Kifley A, Kurrle S, et al. Increasing the uptake of vitamin D supplement use in Australian residential aged care facilities: results from the vitamin D implementation (ViDAus) study[J]. [BMC Geriatrics](#), 2020, 20(1): 383.

[31] Kim H, Park YH, Jung YI, et al. Evaluation of a technology-enhanced integrated care model for frail older persons: protocol of the SPEC study, a stepped-wedge cluster randomized trial in nursing homes[J]. [BMC Geriatrics](#), 2017, 17: 88.

[32] Laborie S, Denis A, Horsch A, et al. Breastfeeding peer counselling for mothers of preterm neonates: protocol of a stepped-wedge cluster randomised controlled trial[J]. [BMJ Open](#), 2020, 10 (1): e032910.

[33] Hooper R, Kasza J, Forbes A. The hunt for efficient, incomplete designs for stepped wedge trials with continuous recruitment and continuous outcome measures[J]. [BMC Medical Research Methodology](#), 2020, 20(1): 279.

[34] Barker D, McElduff P, D'Este C, et al. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available[J]. [BMC Medical Research Methodology](#), 2016, 16: 69.

[35] 钱莎莎, 邢健男, 王璐. 多水平统计模型分析方法及其应用 [J]. [中国公共卫生](#), 2017, 33(9): 1414 – 1416.

[36] Twisk J, Proper K. Evaluation of the results of a randomized controlled trial: how to define changes between baseline and follow-up[J]. [Journal of Clinical Epidemiology](#), 2004, 57(3): 223 – 228.

[37] Kotz D, Spigt M, Arts ICW, et al. The stepped wedge design does not inherently have more power than a cluster randomized controlled trial[J]. [Journal of Clinical Epidemiology](#), 2013, 66(9): 1059 – 1060.

[38] Hemming K, Girling A. The efficiency of stepped wedge vs. cluster randomized trials: stepped wedge studies do not always require a smaller sample size[J]. [Journal of Clinical Epidemiology](#), 2013, 66(12): 1427 – 1428.

[39] Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models[J]. [Statistics in Medicine](#), 2016, 35(13): 2149 – 2166.

[40] Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples[J]. [Trials](#), 2017, 18(1): 101.

[41] Copas AJ, Lewis JJ, Thompson JA, et al. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches[J]. [Trials](#), 2015, 16: 352.

[42] Brown SP, Shoben AB. Information growth for sequential monitoring of clinical trials with a stepped wedge cluster randomized design and unknown intracluster correlation[J]. [Clinical Trials](#), 2020, 17(2): 176 – 183.

[43] Martin J, Taljaard M, Girling A, et al. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials[J]. [BMJ Open](#), 2016, 6 (2): e010166.

[44] Joag K, Ambrosio G, Kestler E, et al. Ethical issues in the design and conduct of stepped-wedge cluster randomized trials in low-resource settings[J]. [Trials](#), 2019, 20(S2): 703.

[45] Hemming K, Taljaard M, McKenzie JE, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration[J]. [BMJ](#), 2018, 363: k1614.

[46] Hemming K, Taljaard M, Grimshaw J. Introducing the new CONSORT extension for stepped-wedge cluster randomised trials[J]. [Trials](#), 2019, 20(1): 68.