

DOI: 10.13288/j.11-2166/r.2016.17.011

方法学与临床评价

阶梯设计方法及其在中医临床研究中的应用

王彬¹, 闫世艳², 马利[✉]

(1. 湖北中医药大学信息工程学院, 湖北省武汉市洪山区黄家湖西路一号, 430065; 2. 中国中医科学院中医临床基础医学研究所)

摘要 阶梯设计是一种特殊的群组随机对照试验,其主要特点是分阶段实施干预,不用设置专门的对照组,既做到了平行随机对照,又避免了伦理学问题。近 10 年来,该方法逐渐被研究者关注,相关理论也趋于完善。介绍该方法在设计、实施和统计分析方面的要点,提出对艾滋病、结核病、肝炎等慢性疾病的中医药治疗及管理等方面的临床研究具有借鉴意义的阶梯设计。

关键词 阶梯设计; 群组随机; 随机对照试验; 临床研究

随机对照试验被认为是评价临床干预有效性的“金标准”^[1]。根据随机的对象,可以将随机对照试验分为个体随机对照试验(RCTs)和群组随机对照试验(CRTs)两种,区别在于前者针对受试者个人进行随机化分配,后者是针对群组的随机化分配,不同群组内受试者的试验条件基本一致^[2]。阶梯设计试验(stepwedgetrials,SWTs)是一种特殊群组随机对照试验,最早出现在冈比亚肝炎干预研究中^[3]。近年来,SWTs 方法使用逐渐增多^[4],以“the stepped wedge design”为关键词检索 PubMed 数据库,检索到从建库至今报道 SWTs 的文献 199 篇,其中 189 篇为 2010 年后发表的文献,文献的主要类型为临床试验(81 篇)、综述(14 篇)和方法学研究(13 篇);研究领域集中在肿瘤(15 篇)、艾滋病(13 篇)、药物毒理性研究(11 篇)、补充和替代医学(10 篇)和伦理学(8 篇)等方面,其中补充和替代医学研究领域无一提及中医药。近年来 SWTs 也已经受到研究者的关注和重视,但国内目前尚无 SWTs 的相关临床研究,仅有一篇文献^[5]简要介绍了 SWTs 的方法。因此,有必要对 SWTs 方法再次进行跟踪研究,以期对中医临床研究提供新的方法学思路。

1 阶梯设计的基本原理与方法

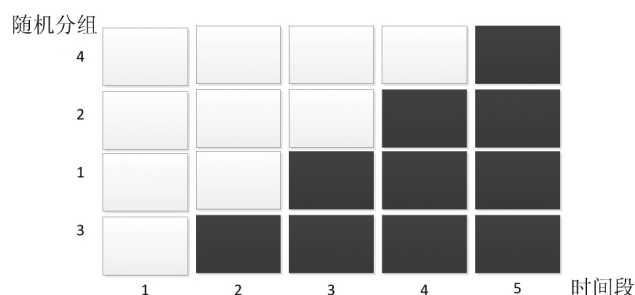
1.1 阶梯设计的基本原理

SWTs 是一种常用于评价“利大于弊”干预措施(诸如疫苗接种、疾病预防、健康教育等)的 CRTs,其基本原理是根据研究目的将研究对象分为若干个小组,并对其进行随机编号;按照时间先后顺序将干预过程划分为不同的阶段。研究开始后,按照事先确定的随机编号顺序给予对应小组干预,已纳入的小组将在研究过程中持续接受干预,而未纳入的小组则保持空白等待状态,直到轮到其接受干预,如此反复直至所有小组均接受干预^[6]。以图 1 为例,将整个试验人群按照干预次数分为 4 个小组,并根据干预时间划分 5 个时间段,通过抽签获得群组随机编号为 3、1、2、4。第 1 个时间段内均不接受干预,后面每一个时间段内均按照编号顺序纳入对应的小组接受干预,直至第 5 个时间段结束时,所有小组都接受了干预。

由此可见,SWTs 设计方法的基本特点是:1) 可以不用设置专门的对照组,由于该方法多用于“利大于弊”的干预措施的研究,因此,这样做可以使更多的人接受干预而避免伦理学问题。2) 研究对象不是同时接受干预,而是按照随机的顺序,分阶段接受干预,最终被全面覆盖,相比于所有组同时接受干预,可以减轻试验在人力、物力、财力方面的压力。

基金项目:国家自然科学基金(81303149);中央级公益性科研院所基本科研业务费专项资金(ZZ070818);湖北省厅级重点项目(D20152003)

* 通讯作者: mary. ht@126.com, (027) 68890154



注：白色方格表示该组群处于对照组；黑色方格表示该组群接受干预，为试验组

图1 阶梯设计试验原理图

1.2 阶梯设计的方法

根据研究对象类型不同，Andrew 等将 SWTs 设计方法^[4]分为以下3种：

1.2.1 固定队列阶梯设计 (closed cohort design)

研究开始前即确定好将要入组的所有研究对象，研究对象将参加整个研究阶段，不会在研究过程中纳入或补充新的研究对象，且不允许研究对象在群组中进行转换。研究者通过对同一研究对象进行多个时点的重复测量，来评价干预措施的有效性 (如图2 A所示)。此方法适用于人员数目相对固定的组群，但不宜用来评价严重疾病或死亡等结局指标，因为退出观察或失访易导致幸存者偏倚。

1.2.2 开放队列阶梯设计 (open cohort design)

试验过程中有部分研究对象退出研究时，可以补充其他符合纳入、排除标准的合格研究对象加入试验。试验中可允许有少数研究对象发生群组间的转换 (如图2 B所示)。由于研究对象撤离而导致数据缺失，研究者普遍采用横断面抽样个体在预先确定时间点上的数据分析感兴趣的结果。研究集群设置较大，每个时间点只有小比例的参与者被抽样，所以参与者通常最多被测量1次。

1.2.3 连续入组短期暴露阶梯设计 (continuous recruitment with short exposure)

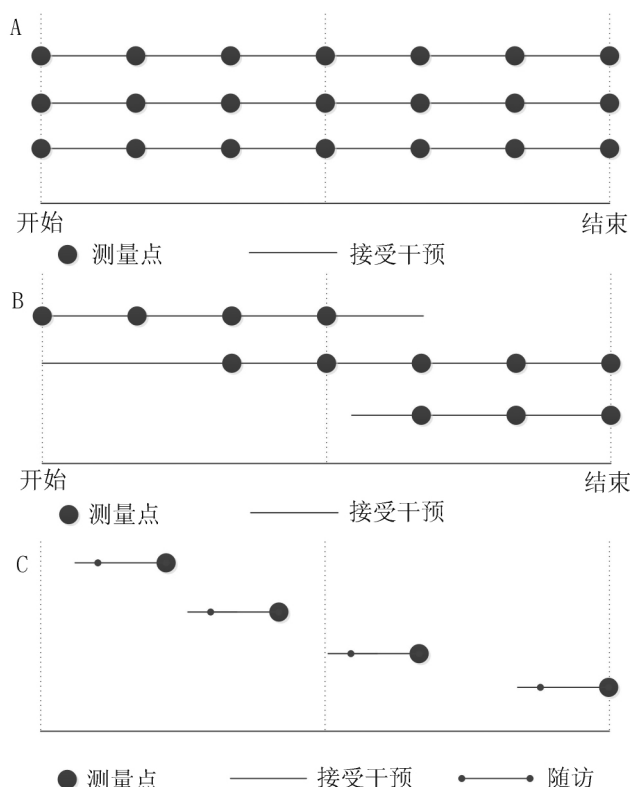
研究开始时，很少或者没有研究对象入组，随着时间推移，更多满足试验条件的研究对象进入试验，并接受短期暴露 (干预)。该种设计下，研究者通常通过后期随访获取数据 (如图2 C所示)。该方法中研究对象仅需接受对照或干预措施中1种，在多数情况下，研究只需要1个测量结果。

2 阶梯设计中的关键问题

2.1 随机化

SWTs 属于 CRTs 的一种特殊形式，其随机化与 CRTs 相同。研究表明，CRTs 中进行随机化时，

若组群数较少，会出现研究对象重要特征分配的不



注：A，固定队列阶梯设计方法图解；B，开放队列阶梯设计方法图解；C，连续入组短期暴露阶梯设计图解

图2 阶梯设计方法示意图

均衡，组群间差异大而导致检验效能降低^[5]。针对该问题，可以采用分层和匹配的方法来解决。分层即分层随机法，先对组群实施分层，然后在不同的层内再进行随机，按照随机分配结果的顺序接受干预。如在赞比亚加强公共卫生 SWTs 研究 (BHOMA)^[6]中，将42个组群按研究对象特征划分为A、B、C三层，每层分别包含7、14和21个组群，在每层对组群随机化后，按照编号顺序依次接受干预。由于随机时考虑了分层因素，因此，统计分析时将分层作为固定效应进行处理。

2.2 步数和步长设计

步数即研究划分的阶段数，步长即每个阶段的持续时间。Hussey 等^[4]在研究中指出，在样本量一定的情况下，干预步数越少，试验测量的次数越少，统计效能就越低。然而在实际研究中，步数太多又会加大后勤压力，增加费用，延长试验时间，给试验参与者增加心理压力。因此，合理的步数制定需要研究者根据研究条件，权衡时间与经济关系，尽可能提高统计效能。此外，一个组群接受干预后，干预起效需要时间，这样给数据测量造成一定延时。因此，在 SWTs 步长设计时，为得到准确

的干预后数据, 步长应大于干预延迟时间^[6]。

2.3 样本量计算

合适的样本量有助于研究者利用最合理的资源获得最可靠的研究结果。SWTs 是对群组进行随机, 其样本量估算可以借鉴 CRTs 的方法, 但是 SWTs 在进行样本量估算时, 还需要考虑步数、群组间的相关性和延迟效应等, 因此, 目前应用较多的是 Hussey 等^[14]提出的样本量计算方法。除该方法外, 有研究者提出了多种基于效应差值的样本量计算方法^[11]。但是这几种方法均是利用 SWTs 设计中横断面测量点上的数据进行样本量估算, 当研究设计是多个时点重复测量时, 这样往往会高估样本量, 而且对于一些特殊设计的 SWTs, 需要对这些样本量估算的基本公式进行修改调整, 而不能直接使用。故此有研究者提出基于模拟的样本量估算方法, 并认为该方法更为可靠和具有普适性^[12]。总之, SWTs 在进行样本量估算时, 要额外考虑其设计的特殊之处。

2.4 结果分析和统计学问题

在统计分析方面, 同样可以借鉴 CRTs 的分析方法。SWTs 结果分析比较方式有同期比较和前后比较两种^[9]。同期比较指分别将不同时间段内接受干预对象与其他组未接受干预对象进行比较; 前后比较则是在达到全面干预之前, 将各时间段内接受干预与未接受干预对象汇总后进行比较, 通常两种比较方式都会采用。另外, 结局指标的时间变化趋势也是常常要考虑的一个因素, 因此, 在前后比较中, 通常要将时间作为一个固定效应因素纳入模型中。此外, 群组内的相关性和受试者的多次重复测量之间的相关性也要考虑。COX 回归模型、Logistic 回归模型以及 Poisson 回归模型等常被使用。

3 实例介绍

实践指导是一种在初级保健中实现基于循证的最佳实践方法, 主要是在某项实践活动中引入一名外部的专业医疗保健人员, 以便帮助初级保健人员顺利应对实践活动实施过程中带来的问题, 确保活动获得数据最佳, 初级保健人员通常是具有管理经验的培训过的护士。2007 年 IDOCC (the Improvement Delivery of Cardiovascular Care) 开展了利用实践指导方法提升患者依从疾病防治指导方针的 SWTs 研究^[13], 疾病包括心脏病、中风、外周血管疾病、肾脏疾病和糖尿病。研究在加拿大安大略省的东部进行, 包含了渥太华和周边的农村社区, 该

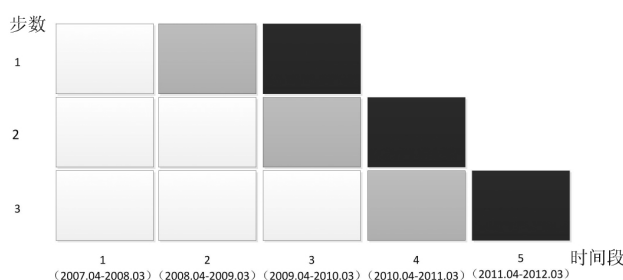
地区人员患慢性疾病机率在加拿大五大湖中最高。该研究采用 SWTs 设计的原因是: 1) 可以大大节省该项大型实践研究的人力、物力、财力; 2) 可以控制时间的影响; 3) 可以保证该研究中所有的受试者在研究结束的时候都接受干预。

研究以进行医疗实践点为基本单元, 是一种开放队列阶梯设计。整个研究分三步, 根据地理信息系统映射技术, 将研究区域分割为 9 块, 并根据研究步数, 按照地理位置将每个地区再划分为三层 (西、中、东), 对每层进行随机化, 使其接受三步中的任意一步, 研究示意图如图 3 所示。研究采用两组比较的平行对照试验的样本量计算方法, 在前期研究基础上, 经过家庭医生和心脏病专家共同商量, 假定组群内相关系数为 0.18、共同标准差为 18%, 并考虑 20% 脱落的情况下, 每步纳入 30 个医疗实践点, 每个医疗实践点平均包含 66 个患者。由于 SWTs 在研究期间进行多次重复测量, 其统计效率应高于两组的平行设计研究, 样本量得以保证。

研究干预通过指导者与患者进行定期会议或走访进行, 每步分为两阶段 (密集阶段和持续阶段), 时长各 1 年, 密集阶段干预次数多达 13 次 (4 周 1 次), 持续阶段干预次数为 4~6 次 (8~12 周 1 次), 每次干预前后都要收集相关数据, 如心血管疾病情况、糖尿病或肾病情况、吸烟状态、血压、体重、血脂等, 并通过这些数据计算训练等级, 训练等级越高表示患者的依从性越好。结果分析采用 Hussey 等^[14]提出的一般线性混合回归模型分析, 其中时间、处理阶段、区域作为固定效应, 把实践点作为随机效应因素, 采用一个复合对称的协方差结构来解释同一受试者多次重复测量间的相关性。研究结果说明 IDOCC 并没有改善患者依从最佳预防措施的情况, 但研究很好地实现了在大区域内大样本量的试验研究, 有效解决了试验经费问题。

4 在中医临床研究中的应用前景

对于中医领域的临床研究来说, 艾滋病、肝炎、结核病以及高血压病等慢性疾病的干预和管理可以借鉴该方法。一方面, 中医的治疗方案多来自于长期的临床实际应用, 其有效性已经得到了初步的检验, 有利于阶梯设计时所有受试者都接受该干预。另一方面, 对于到中医临床研究机构就医的患者来说, 若采用传统的 RCT 设计, 患者往往难以



注：白色格表示未接受干预处于对照组；灰色格表示接受干预处于密集期；黑色格表示接受干预处于持续期

图3 IDOCC 阶梯试验设计图

接受将其随机分配到安慰剂组或西医组。此外，在艾滋病临床研究中，由于我国艾滋病分布具有显著的地区聚集性，因此很容易实施群组随机，即将某个村子的患者作为群组，实施群组随机化，既保证了随机原则，同时也便于实施，有利于保持盲态、减少沾染和失访。在国外目前进行的 SWTs 研究中，用于艾滋病临床研究的较多，就目前我国中医治疗艾滋病的临床研究现状来看，基本均是在西药抗病毒治疗的基础上加载中医治疗方案，这种情况下，无论是伦理学上还是试验的实施上，与安慰剂对照相比，采用 SWTs 更为合适。SWTs 对于肝炎、结核病以及肿瘤等疾病的临床研究也同样适用。这些慢性疾病均是中医治疗的优势病种，且临床研究往往周期较长，多随访 2~3 年，若采用安慰剂对照，随着随访时间的延长，必然会导致较大比例的脱落。采用 SWTs，可以很好地避免这个问题，并且可以大大降低成本。

目前我国无论西医还是中医临床研究，均无 SWTs 方面的临床应用实例。结合该研究设计的特点和中医临床研究实际情况，我们认为，该方法值得借鉴和推广，在下一步的工作中，本课题组将设计 SWTs 方面的临床研究。

5 讨论

SWTs 具有各群组分阶段随机接受干预的特点，因此，可以做到最大限度地使受试者接受较为有益的处理，保持了试验研究中的随机和对照原则，避免了伦理学问题。该方法适用于评价“利大于弊”的治疗措施，或现有资源有限而无法全面开展研究的情况。在公共卫生领域中，当研究者希望在一定范围内全面推行某项措施，同时又想评价该措施的效果时，该方法也非常适用^[3]。

当然，SWTs 也存在不足的地方，如无法做到双盲，只能对结局评价者设盲^[4]；可能存在生存者偏倚^[5]和阶梯设计效应，即在样本量相同的情

况下，相比于 RCTs、SWTs 获得的统计量偏小，导致犯 II 类错误的概率增加^[3]。除此之外，SWTs 还存在治疗效应延迟问题，即实施干预后的效果在其后一个或多个时间段随访中被检测出来^[4]，会导致不同时间点测量的数据偏小，低估药物疗效。研究者在设计 SWTs 方案时，要充分考虑干预治疗的延迟效应，设定足够长的时间段（步长）^[4]。

SWTs 为中医药临床科研提供了新的设计方法，但其设计、实施和统计分析都具有较高的复杂性，需要研究者在设计和应用 SWTs 时，更加严谨地计划和实施。

参考文献

- [1] SOLOMON MJ, MCLEOD RS. Surgery and the randomised controlled trial: past, present and future [J]. Med J Aust, 1998, 169(7): 380-383.
- [2] BAO G, COPAS A, AMBLER G, et al. Sample size calculation for a stepped wedge trial [J]. Trials, 2015, 16: 354. doi: 10.1186/s13063-015-0840-9.
- [3] The Gambia Hepatitis Study Group. The Gambia Hepatitis Intervention Study [J]. Cancer Res, 1987, 47(21): 5782-5787.
- [4] 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. doi: 10.1186/s13063-015-0839-2.
- [5] 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.
- [6] BROWN CA, LILFORD RJ. The stepped wedge trial design: a systematic review [J]. BMC Med Res Methodol, 2006, 6: 54. doi: 10.1186/1471-2288-6-54
- [7] 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. doi: 10.1186/s13063-015-0842-7.
- [8] CAMPBELL MK, PIAGGIO G, ELBOURNE DR, et al. Consort 2010 statement: extension to cluster randomised trials [J]. BMJ, 2012, 345: e5661. doi: 10.1136/bmj.e5661.
- [9] STRINGER JS, CHISEMBELE-TAYLOR A, CHIBWESHA CJ, et al. Protocol-driven primary care and community linkages to improve population health in rural Zambia: the Better Health Outcomes through Mentoring and Assessment (BHOMA) project [J]. BMC Health Serv Res, 2013, 13 (Suppl 2): S7. doi: 10.1186/1472-6963-13-S2-S7.
- [10] HUSSEY MA, HUGHES JP. Design and analysis of stepped wedge cluster randomized trials [J]. Contemp Clin Trials, 2007, 28(2): 182-191.
- [11] WOERTMAN W, DE HOOP E, MOERBEEK M, et al.

- Stepped wedge designs could reduce the required sample size in cluster randomized trials [1]. J Clin Epidemiol, 2013, 66(7): 752-758.
- [2] DIMAIRO M, BRADBURN M, WALTERS SJ. Sample size determination through power simulation; practical lessons from a stepped wedge cluster randomised trial (SW CRT) [J]. Trials, 2011, 12(Suppl 1): A26.
- [3] LIDDY C, HOGG W, SINGH J, et al. A real-world stepped wedge cluster randomized trial of practice facilitation to improve cardiovascular care [J]. Implement Sci, 2015, 10(1): 150. doi: 10.1186/s13012-015-0341-y.
- [4] HARGREAVES JR, COPAS AJ, BEARD E, et al. Five questions to consider before conducting a stepped wedge trial [J]. Trials, 2015, 16(1): 350. doi: 10.1186/s13063-015-0841-8.
- [5] MOULTON LH, GOLUB JE, DUROVNI B, et al. Statistical design of THRio: a phased implementation clinic-randomized study of a tuberculosis preventive therapy intervention [J]. Clin Trials, 2007, 4(2): 190-199.
- [6] STERN A, MITSAKAKIS N, PAULDEN M, et al. Pressure ulcer multidisciplinary teams via telemedicine: a pragmatic cluster randomized stepped wedge trial in long term care [J]. BMC Health Serv Res, 2014, 14(1): 83. doi: 10.1186/1472-6963-14-83.

Stepped Wedge Design Methodology and Its Application in Chinese Medicine Clinical Trials

WANG Bin¹, YAN Shiyan², MA Li¹

(1. College of Information Engineering, Hubei University of Chinese Medicine, Wuhan, 430065; 2. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences)

ABSTRACT Stepped wedge design is a special group randomized controlled trial, and its main feature is phased implementation of intervention without setting specific controlled group. This methodology could achieve parallel randomized controlling, and also avoid ethical problems. In the recent decade, this design method is attracting researchers' concern, and the relative theory is becoming complete. This article introduces the key points in designing, implementing and statistical analysis, and proposes the stepped wedge design which could be used in Chinese medicine treating and managing chronic diseases, including acquired immune deficiency syndrome, tuberculosis, and hepatitis.

Keywords stepped wedge design; group randomization; randomized controlled trials; clinical research

(收稿日期: 2016-03-25; 修回日期: 2016-04-14)

编辑: 柴倩云