

阶梯设计样本量估算方法研究进展及展望*

中山大学公共卫生学院医学统计学系(510080) 张沂 郑萌 郭婷婷 王琦 陈雯[△]

【摘要】 目的 对现有阶梯设计样本量估算方法进行综述,为在阶梯设计中进行样本量估算及开发新的估算方法提供依据。**方法** 在 PubMed 和 Embase 中检索自 1986 年至今发表的阶梯设计样本量估算相关文献,共计 32 篇。按照基于设计效应推导和基于数据模拟两种估算思路对现有样本量估算方法进行分类,归纳总结各方法适用的设计特征并分析待解决的问题。**结果** 现有的 11 种阶梯设计样本量估算方法主要适用于连续型结局变量下的标准阶梯设计,适用于非连续型结局变量和开放队列的估算方法较少。对于近年来快速发展出的阶梯设计的亚型,样本量估算研究相对滞后。**结论** 阶梯设计的样本量估算方法滞后于设计本身的快速发展,为适应设计类型的快速发展,需要开展相应样本量估算研究。

【关键词】 阶梯设计 样本量 设计效应 数据模拟

【中图分类号】 R195.1 **【文献标识码】** A

DOI 10.11783/j.issn.1002-3674.2025.01.030

阶梯设计群随机试验 (stepped wedge cluster randomized trails, SW-CRT), 又称“阶梯设计” (stepped wedge design, SWD), 是群随机对照试验 (cluster randomized trails, CRT) 的一个变体, 1986 年在冈比亚肝炎干预研究中第一次得到实践^[1]。在 SWD 中, 所有的群在试验开始时都处于对照状态, 称为基线期, 然后按照随机分配的顺序依次从对照状态转换至干预状态 (单向转换)^[2], 在最后一个时期, 所有群都接受干预, 称为完成期 (图 1)。单向转换的特点是阶梯设计与平行 CRT 以及双向群交叉设计的区别之处, 在平行 CRT 中, 群的干预或对照状态在随机分配后便不再改变; 双向群交叉设计则具有双向转换特点, 即可以从对照状态转换至干预状态, 也可以从干预状态转换为对照状态^[3]。

前后对照, 除非群内相关系数非常小 (如小于 0.02^[3]), 且群的规模也不大时, 相比于 CRT, SWD 会是更适宜的设计类型; SWD 还可以解决试验资源有限, 无法同时开展大规模 CRT 的问题^[4-6]。

近年来随着实施科学的兴起和发展, SWD 越来越地被用于医疗卫生服务和公共卫生政策的效果评价研究^[7-8]。然而, 系统综述显示大多数 SWD 没有使用与研究设计相匹配的样本量估算方法^[9]。SWD 样本量估算方法研究也滞后于研究设计的发展^[10]。本文由两名作者 (张沂和郑萌) 检索 PubMed 和 Embase 数据库中收录的 1986 年至今发表的文献 (检索式及文献筛选流程图见图 2), 对直接能得出样本量估算结果或程序的方法, 按照基于设计效应推导和数据模拟的两种思路总结当前 SWD 样本量估算方法的研究进展, 分析尚待解决的研究问题, 可为 SWD 的样本量估算及开发更适宜的样本量估算方法提供依据和指导。

阶梯设计简介

在标准的 SWD 中, 共有 g 个序列 (sequences) 和 $T (T = g + 1)$ 个长度相等的时期 (periods), 每个序列内有相等数量的群 (clusters), 且每个群大小相等。

基于抽样结构, SWD 可分为横断面 SWD (cross-sectional SWD) 和队列 SWD (cohort SWD)。横断面 SWD 在每个时期, 从群中随机抽取一定数量的个体纳入试验, 因此每次观测的个体并不固定。队列 SWD 则要求在研究开始前就确定好所有入组的个体, 对相同的个体进行多次测量。设研究中有 I 个群, 每个群大小为 m , 每个群在每个时期的大小为 n , 则横断面 SWD 中 $m = n \times T$, 队列 SWD 中 $m = n$ 。此外, 根据被纳入的个体是否完整地参与研究, 队列 SWD 可再分为固定队列 SWD (closed cohort SWD) 和开放队列

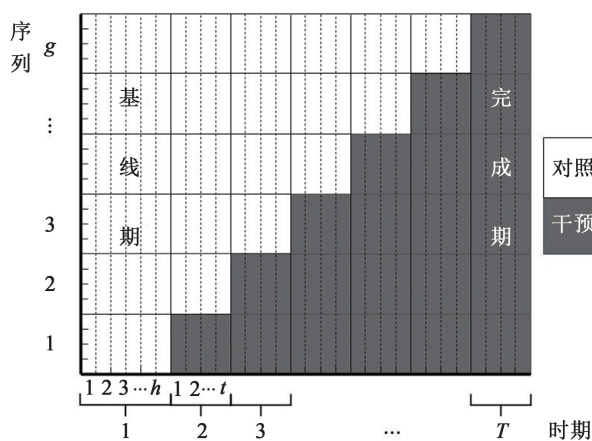


图 1 SWD 示意图

在该设计下, 所有参与试验的个体都可接受干预, 因此它尤其适用于评价“获益风险比高”的干预, 可很好地规避伦理问题。此外, SWD 中每个群都可作自身

* 基金项目: 广东省自然科学基金 (2022B1515020094); 美国中华医学基金会 (20-389); 国家自然科学基金 (72074229)
[△] 通信作者: 陈雯, E-mail: chenw43@mail.sysu.edu.cn

SWD(open cohort SWD) [10-12]。

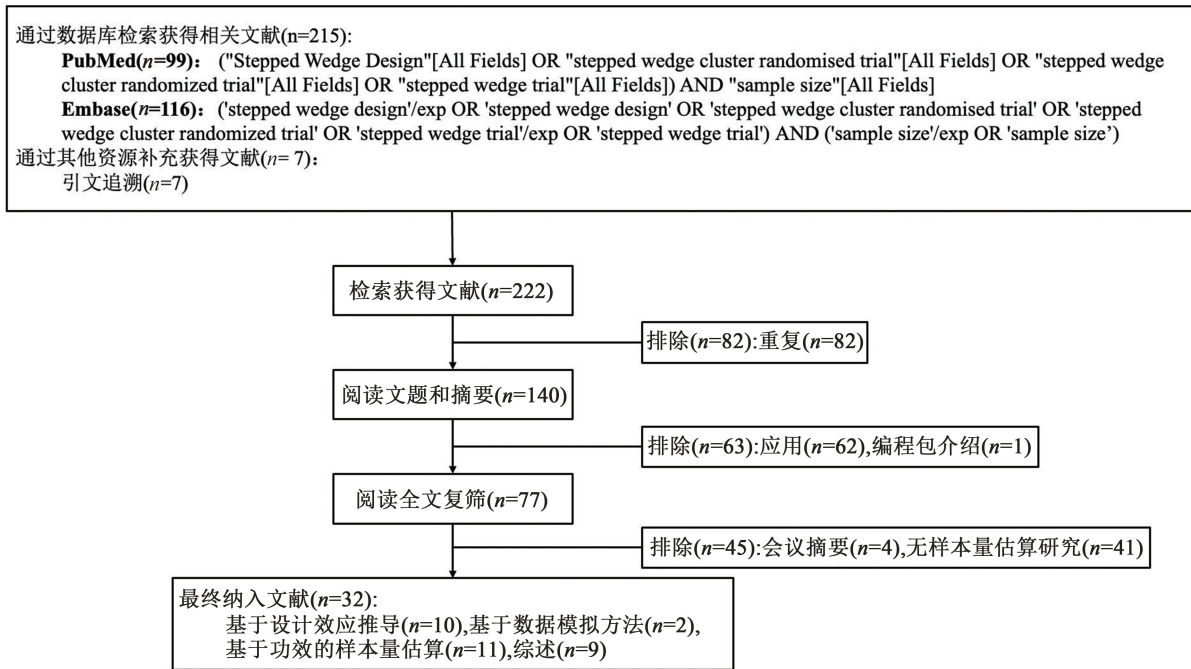


图2 文献检索及筛选流程图

标准的 SWD 在实际使用时存在局限,比如只能用来评估一个处理因素的效果,无法进行多组比较。为解决实际问题,近年来 SWD 快速发展出多种亚型,如分批次^[13]、含有两个以上聚类层级^[14-15]、含有多个干预项^[16]的 SWD 等。

基于设计效应推导的 SWD 样本量估算

在 SWD 中,样本量为整个试验过程中参与研究的受试者总数,记为 N_{swd} 。基于设计效应 (design effect, deff) 的样本量估算首先是在不同设计类型及前提条件下推导出设计效应,再通过下式估算总样本量:

$$N_{swd} = N_u \times deff \quad (1)$$

其中, $deff$ 为 SWD 试验中处理效应估计值方差与相同样本量下个体随机试验 (individually randomized trial) 中处理效应估计值方差的比值, N_u 是个体随机试验达到规定检验功效所需的样本量^[17]。下文将根据不同的设计特征归纳相应的设计效应。

1. 横断面 SWD 的设计效应推导

(1) SWD 设计效应推导的基础——HH 方法

Hussey 和 Hughes^[2] 于 2007 年率先提出了横断面 SWD 检验功效的计算公式,简称为 HH 方法。之后基于设计效应推导的 SWD 样本量估算方法都以此为基础。

HH 方法的前提条件为:①存在时间效应且为固定效应,②不存在群×时间和群×个体的交互效应,③无论处于对照还是干预时期,来自同一群的两个个体间的相关性相同,④ n 恒定不变。该方法中检验功效的计算基于线性混合效应模型 (linear mixed effects

model, LMM):

$$Y_{ijk} = \mu_{ij} + e_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta + e_{ijk} \quad (2)$$

其中, Y_{ijk} 是群 i 中个体 k 在时期 j 处的响应变量 ($i = 1, 2, \dots, I; j = 1, 2, \dots, T; k = 1, 2, \dots, n$), 总方差为 σ_y^2 , μ 是截距项, α_i 是群 i 的随机效应, $\alpha_i \sim N(0, \tau^2)$, β_j 是时期 j 的时间效应, X_{ij} 是设计矩阵,其元素为群 i 在时间点 j 干预与否的指示变量 (1 = 干预; 0 = 对照), θ 是处理效应, e_{ijk} 是个体随机效应, $e_{ijk} \sim N(0, \sigma_e^2)$ 。

检验水准 α 下的检验功效为:

$$power = \Phi\left(\frac{\hat{\theta}}{\sqrt{Var(\hat{\theta})}} - z_{1-\alpha/2}\right) \quad (3)$$

其中, Φ 是累积标准正态分布函数, $z_{1-\alpha/2}$ 是标准正态分布函数的第 100(1- $\alpha/2$) 分位数, $\hat{\theta}$ 为处理效应估计值,其方差为:

$$Var(\hat{\theta}) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{(IU - W)\sigma^2 + (U^2 + ITU - TW - IV)\tau^2} \quad (4)$$

其中, $U = \sum_{ij} X_{ij}$, $W = \sum_j (\sum_i X_{ij})^2$, $V = \sum_i (\sum_j X_{ij})^2$, $\sigma^2 = \frac{\sigma_e^2}{n}$ 。 I 为群个数, T 为时期数。

(2) 基于 HH 方法的设计效应推导

在 Woertman 等人^[18] 的研究中 (图 1), T 个时期被分割为 $h+gt$ 个测量时点, h 是基线期包含的测量时点数, t 是之后每个时期包含的测量时点数。除满足 HH 前提条件外, Woertman 等人还假设每个序列内有相等数量的群且 t 是固定值。Woertman 等人基于公式 (4) 推导出横断面 SWD 的设计效应:

$$deff_1 = (h+gt) \times \frac{1+\rho(gtn+hn-1)}{1+\rho(\frac{1}{2}gtn+hn-1)} \times \frac{3(1-\rho)}{2t(g-\frac{1}{g})} \quad (5)$$

其中, ρ 是群内相关系数 (intraclass correlation coefficient, ICC), 表示同一群中个体间的相关性, 其他符号定义同前。

若 $h=t=1$, 则设计效应简化为^[19]:

$$deff_2 = (1+g) \times \frac{1+\rho(gn+n-1)}{1+\rho(\frac{gn}{2}+n-1)} \times \frac{3(1-\rho)}{2(g-\frac{1}{g})} \quad (6)$$

对于不满足 HH 前提条件的情况, 可考虑采用以下两种样本量估算方法。

当 HH 前提条件①不成立, 即不存在时间效应^[20], 其他条件同前时, Zhou 等人^[21] 推导设计效应为:

$$deff_3 = \frac{1+(ng-1)\rho}{1+\frac{2}{3}n(g+1)\frac{\rho}{1-\rho}} \quad (7)$$

当 HH 前提条件④不成立, 即 n 不为固定值, 其他条件同前时, Harrison 等^[22] 借助群大小的变异系数推导出的设计效应为:

$$deff_4 = \frac{3(T-1)(1-\rho)[1+(T\bar{n}-1)\rho]}{(T-2)\{2(1-\rho)+(T+1)\bar{n}\rho - \frac{T+1}{IT}\kappa^2[1+(T\bar{n}-1)\rho]\}} \quad (8)$$

其中, n_i 为每个时期各个群的大小, $\kappa^2 = \frac{I^2}{(\sum_{i=1}^I n_i)^2(I-1)} \sum_{i=1}^I (n_i - \bar{n})^2$ 为群大小的变异系数, $\bar{n} = \frac{\sum_{i=1}^I n_i}{I}$, 其他符号定义同前。

此外, 当基线期测量时点数 h 和完成期测量时点数 a 都变化时, Thompson 等人^[5] 推导出的设计效应为:

$$deff_5 = \frac{3[1+\rho(m-1)]g(g-1)(1-R)}{2(g+1)[1-(h+a)n/m]\{g-0.5gR[1-(h+a)n/m]-1\}} \quad (9)$$

其中, $R = \frac{mp}{1+(m-1)\rho}$ 为群平均相关性 (cluster-mean correlation, CMC), 其他符号定义同前。

2. 队列 SWD 的设计效应推导

如果将横断面 SWD 的样本量估算方法直接用于队列 SWD, 则会高估所需样本量。在队列 SWD 中, 研究者在早期横断面设计的基础上, 引入时期内 ICC (within-period ICC, wp-ICC) 和时期间 ICC (between-period ICC, bp-ICC), 前者表示同一群中两个个体在同一时期内的相关性, 后者表示同一群中两个个体在不同时期内的相关性, 并且允许二者不相等。bp-ICC 与 wp-ICC 的比值称为群自相关系数 (cluster au-

tocorrelation coefficient, CAC), 通常小于 1。此外, 在队列设计中, 需要通过额外相关性来体现同一个体在不同时期存在重复测量, 将其量化为个体自相关系数 (individual autocorrelation coefficient, IAC)^[23]。为方便计算, 设计效应的推导中通常假定 IAC 和 CAC 是固定值。

(1) 固定队列 SWD 的设计效应推导

Girling 等人^[24] 在公式 (2) 的基础上, 通过加入群 \times 时间的交互效应项 $(ct)_{ij}$ 和群内随机效应 s_k 、 $(st)_{kj}$ (方差分别为 $\eta_{ct}\sigma_y^2$ 、 $\eta_s\sigma_y^2$ 和 $\eta_{st}\sigma_y^2$) 来处理 IAC 和 CAC, 推导出固定队列 SWD 的设计效应:

$$deff_6 = \frac{n\eta_{ct} + \eta_{st}}{\frac{2}{3}(1-\frac{1}{g}) - \frac{1}{3}(1-\frac{2}{g+1})R_G} \quad (10)$$

其中, $R_G = \frac{T\rho'}{1+(T-1)\rho'}$, $\frac{\rho'}{1-\rho'} = \frac{n\tau^2/\sigma_y^2 + \eta_s}{n\eta_{ct} + \eta_{st}}$, 其他符号定义同前。

Hooper 等人^[25] 则直接纳入 IAC (ϑ) 和 CAC (π) 推导固定队列 SWD 的设计效应为:

$$deff_7 = \frac{3g(1-R)(1+gR)}{(g^2-1)(2+gR)} = [1+(n-1)\rho] \quad (11)$$

其中, $R = \frac{n\rho\pi + (1-\rho)\vartheta}{1+(n-1)\rho}$ 为 CMC, 其他符号定义同前。

上述公式 (5) 到公式 (11) 都是基于 LMM 进行的推导, 仅适用于连续型结局变量的样本量估算。Li 等人^[26] 基于广义估计方程 (generalized estimating equations, GEE) 推导出固定队列 SWD 的设计效应为:

$$deff_8 = \frac{3}{2t(g-1/g)} \left[\frac{(h+gt)\lambda_1\lambda_2}{(gt/2)\lambda_1 + (h+gt/2)\lambda_2} \right] \quad (12)$$

其中, $\begin{cases} \lambda_1 = 1+(n-1)(\alpha_0 - \alpha_1) - \alpha_2 \\ \lambda_2 = 1+(n-1)\alpha_0 + g(n-1)\alpha_1 + g\alpha_2 \end{cases}$, α_0 为 wp-ICC, α_1 为 dp-OCC, α_2 为个体内 ICC (within-individual ICC, wi-ICC) 其他符号定义同前。

队列研究中易出现数据缺失、没有遵循标准的阶梯过程实施干预、个体间和个体内的相关结构不固定等问题。在群大小 m 已知且为固定值时, Wang 等人^[27] 基于 GEE 推出了适用于上述更灵活的实际场景样本量的估计:

$$deff_9 = \frac{\sum_{s=1}^g p_s (v_s - \bar{u})' [\tilde{\delta}^\circ \Omega + (m-1)diag(\delta)\Psi diag(\delta)] (v_s - \bar{u})}{n [\sum_{i=1}^T \delta_i \bar{u}_i (1 - \bar{u}_i)]^2} \quad (13)$$

其中, $p_s (s=1, \dots, g)$ 表示每个序列在每个时期接受干预的概率, 在标准 SWD 中, $p_s = \frac{1}{g}$; $v_s (s=1, \dots, g)$ 对应标准 SWD 设计矩阵中的各行向量, $\bar{u} = (\bar{u}_1, \dots,$

\bar{u}_T) 描述每个时期个体接受干预的概率, δ_t 是 t 时刻缺失数据的概率, $\delta = (\delta_1, \dots, \delta_T)'$, $\tilde{\delta}$ 是一个 T 维方阵, 其 (t, t') 位置处的元素为 $\delta_{tt'} = \delta_t \delta_{t'}$, 运算符 \circ 是 Hadamard 乘积, Ω 是个体内相关性矩阵, Ψ 是同一群内个体间相关性矩阵。总样本量为 $N_u \times deff_9 \times m$ 。

(2) 开放队列 SWD 的设计效应推导

在公式(11)的基础上, Kasza 等人^[28] 假设每个序列中只有一个群, 即 $I=g$ 。推导出开放队列 SWD 的设计效应:

$$deff_{10} = \frac{1}{4} \times \frac{g^2(1-R)[1+(T-1)R]}{gU-W+[U^2+g(T-1)U-(T-1)W-gV]R} (1+(n-1)\rho) \quad (14)$$

其中, $R = \frac{\eta\rho\pi+(1-\rho)\vartheta(1-\chi)}{1+(n-1)\rho}$ 为 CMC, χ 为脱失率, 是取值范围为 $[0, 1]$ 的常量, 固定队列中 $\chi=0$, 横

断面中 $\chi=1$ 。其他符号定义同前。

基于数据模拟的 SWD 样本量估算

实际研究中 SWD 有许多复杂的扩展设计, 无法或很难推导设计效应。Baio 等人^[29] 提出通过数据模拟的方法估算样本量。此方法通过设定模拟数据生成过程, 做大量的模拟计算, 通过分析模型对得到的“虚拟试验”进行分析。

最佳样本量设置为: 在给定的检验水准上, 能正确地推断出处理效应估计值有统计学意义的模拟试验的次数与总试验次数的比值 \geq 所需检验功效时, 参与研究的个体总数的最小值。具体估算步骤如图 3 所示。具体模拟过程本文不做演示。此方法具有灵活性高的优点, 同时也存在工作量大、耗时长、效率低的问题。

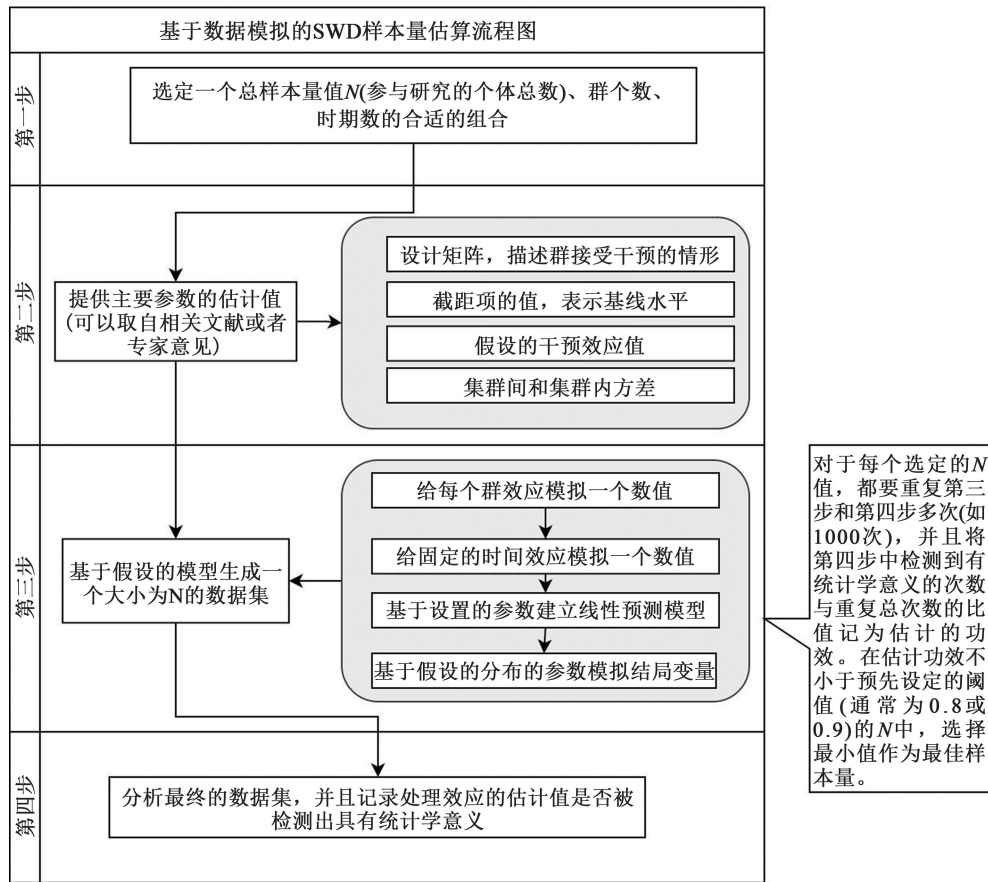


图 3 基于数据模拟的 SWD 最佳样本量估算示意

Zhan 等人^[30] 还通过此思路, 基于贝叶斯方法纳入时间效应的先验信息进行统计分析, 发现可减少所需样本量。

讨论

自 1986 年 SWD 得到实践之后的很长一段时间内, 此设计都缺少相应的样本量估算方法。2007 年, Hussey 和 Hughes^[2] 提出横断面 SWD 检验功效的计

算公式, 为基于设计效应的 SWD 样本量推导提供了方法学基础。在此之后许多学者在标准的 SWD 下基于连续型结局变量, 对如何通过设计效应的推导来估计样本量进行研究。本文综述了现有的 11 种基于设计效应或基于数据模拟的 SWD 样本量估算方法 (Zhan 等人^[30] 使用的研究方法思路同 Baio 等人^[29] 归纳的方法, 故未纳入总结表) 并总结了它们适用的各设计特征 (表 1), 便于研究者基于实际的研究设计

选择适宜的方法。

从设计类型角度,目前的方法主要解决了横断面和固定队列 SWD 样本量估算的问题,适用于开放队列研究较少。多数队列 SWD 样本量估算方法同样适用于横断面 SWD,只需做个别参数的调整。队列 SWD 的样本量估算方法都不满足 HH 方法的假设②和③,具体而言,Girling^[24]、Hooper^[25]和 Li^[26]等人提出的方法都是对于固定队列 SWD 的样本量估算,而 Kasza 等人^[28]的方法是对于开放队列 SWD 的样本量估算。在固定队列 SWD 的三种样本量估计方法中,Girling^[24]和 Hooper^[25]等人的方法是基于 LMM,只适用于连续型结局变量,而 Li^[26]等人的方法基于 GEE 模型,同时适用于连续型和非连续型结局变量。对二分类结局、离散结局的 SWD,HH 两位学者提出的方法是做正态近似推导。但是理论上,二分类和离散结局 SWD 中处理项效应估计值的方差取决于假定的处

理项效应和时间效应的大小^[23],连续型结局变量 SWD 中则不然,所以正态近似的推导方法是不合理的。有研究者基于广义线性混合效应模型进行最大似然估计得到处理项效应的估计值及其方差,或是基于边际模型和 GEE 进行推导,得到基于模型的处理项效应检验功效,并基于此反推需要的样本量,如 Li 等^[26]、Wang 等^[27]和 Xia 等^[31]的研究。其次,相较于标准 SWD 要求各个时期长度相等,Woertman 等^[18]、Thompson 等^[5]和 Li 等^[26]的方法可用于基线期和完成期包含的测量时点数不等于其后各时期的情况。Wang 等^[27]的方法还适用于数据缺失和没有遵循标准的阶梯过程实施干预的情况。总而言之,基于设计效应推导的样本量估算方法随着前提条件和具体设置的变化已不断改进,可有针对性地解决更多问题,但灵活性仍不如基于数据模拟的方法。

表 1 阶梯设计样本量估算方法的使用情景

	设计类型			HH 前提条件*		使用模型		时间长度	
	横断面	固定队列	开放队列	满足	不满足	LMM**	GEE**	相等	不相等
Woertman 等 ^[18] (2013)									
Hemming 等 ^[20] (2016)									
Zhou 等 ^[21] (2017)					①				
Harrison 等 ^[22] (2020)					④				
Thompson 等 ^[5] (2017)									
Girling 等 ^[24] (2016)					②③				
Hooper 等 ^[25] (2016)					②③				
Li 等 ^[26] (2018)					②③				
Wang 等 ^[27] (2021)					②③				
Kasza 等 ^[28] (2016)					②③				
Baio 等 ^[29] (2015)					①②③④				

注:灰色方格表示各方法适用情景。* HH 方法(Hussey 和 Hughes 方法)前提条件:①存在时间效应且为固定效应,②不存在群×时间和群×个体的交互效应,③无论处于对照还是干预时期,来自同一群的两个个体间的相关性相同,④每个时期每个群中的个体数恒定。** LMM 为线性混合效应模型,GEE 为广义估计方程。

通过模拟数据能够在各种复杂情境下估算样本量,但也存在工作量大,效率低的问题;并且此方法的正确性依赖于对相关参数(如 ICC 和 CAC 等)的合理设定。所以,此方法只能为复杂设计提供一种研究的思路和方法框架,基于模拟试验设计的场景得到的结果,可推广性需进一步研究讨论。同时,也亟待开展研究探讨相关性参数取值变化对样本量估算的影响。

目前有较多软件包可以实现阶梯设计的样本量估算,如 R 中的 swCRTdesign 包和 SteppedPower 包,SAS 中的 %swdpwr 包,Stata 中的 steppedwedge 包和 power swgee 包,Ouyang 等人^[23]曾综述了现有的相关编程包,并评价了优劣,研究者可根据实际研究的设计特征进行样本量估算方法的选择。

近年来为解决实际问题,SWD 出现了各种亚型,

如 Teerenstra 等^[14]使用多层级 SWD 来解决群以上层级的聚集性问题,适用于存在三个及以上水平的数据;Kotwal 等^[32]采用多处理 SWD 在一个设计中同时评价多种干预,提高了设计的效率;Kasza 等^[13]通过分批次 SWD 实现群的分批招募并逐步开展多个标准 SWD,可以分阶段扩大研究现场。但这些亚型的样本量估算研究进展缓慢,目前的成果多为在连续结局变量下基于功效反推需要的样本量,无直接的设计效应推导,也无离散结局变量的相关研究。

在各亚型设计的研究中直接使用现有的设计效应公式成果,不去考虑相关性结构、参数估计方法等方面的变化,很可能导致功效的损失,无法保证后续数据分析工作乃至整个试验设计的科学性。为适应 SWD 的快速发展,满足实际研究的需求,研究者应该同步推进

SWD 各亚型样本量估算的研究,为更高效和合理地开展研究提供支持。

参 考 文 献

- [1] The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group[J].Cancer Res,1987, 47(21):5782-5787.
- [2] Hussey M, Hughes J. Design and analysis of stepped wedge cluster randomized trials[J].Contemp Clin Trials,2007, 28(2):91-182.
- [3] Wellek S, Donner-Banzhoff N, König J, et al. Planning and Analysis of Trials Using a Stepped Wedge Design[J].Dtsch Arztebl Int, 2019, 116(26):453-458.
- [4] 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].J Clin Epidemiol, 2013, 66(12):1427-1428.
- [5] Thompson J, Fielding K, Hargreaves J, et al. The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs[J].Clin Trials,2017, 14(6):639-647.
- [6] Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? [J].Int J Epidemiol,2020, 49(3):1043-1052.
- [7] Beard E, Lewis J, Copas A, et al. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014[J].Trials,2015, 16:353.
- [8] 司华新, 乔晓霞, 金雅茹, 等. 阶梯设计试验研究方法及报告规范[J].中国公共卫生,2022, 38(4):503-509.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Copas A, Lewis J, Thompson J, et al. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches[J].Trials,2015, 16:352.
- [13] Kasza J, Bowden R, Hooper R, et al. The batched stepped wedge design: A design robust to delays in cluster recruitment [J].Stat Med,2022, 41(18):3627-3641.
- [14] Teerenstra S, Taljaard M, Haenen A, et al. Sample size calculation for stepped-wedge cluster-randomized trials with more than two levels of clustering[J].Clin Trials,2019, 16(3):225-236.
- [15] Davis-Plourde K, Taljaard M, Li F. Sample size considerations for stepped wedge designs with subclusters [J].Biometrics, 2023, 79(1):98-112.
- [16] Sundin P, Crespi C. Power analysis for stepped wedge trials with multiple interventions[J].Stat Med,2022, 41(8):1498-1512.
- [17] 李莉. 统计学原理与应用[M].南京:南京大学出版社,2019(8):354.
- [18] Woertman W, deHoop E, Moerbeek M, et al. Stepped wedge designs could reduce the required sample size in cluster randomized trials[J].J Clin Epidemiol,2013, 66(7):752-758.
- [19] Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach [J].J Clin Epidemiol,2016, 69:137-146.
- [20] Liao X, Zhou X, Spiegelman D. A note on "Design and analysis of stepped wedge cluster randomized trials" [J].Contemp Clin Trials, 2015, 45:338-339.
- [21] Zhou X, Liao X, Spiegelman D. "Cross-sectional" stepped wedge designs always reduce the required sample size when there is no time effect[J].J Clin Epidemiol,2017, 83:108-109.
- [22] Harrison L, Chen T, Wang R. Power calculation for cross-sectional stepped wedge cluster randomized trials with variable cluster sizes [J].Biometrics,2020, 76(3):951-962.
- [23] Ouyang Y, Li F, Preisser J, et al. Sample size calculators for planning stepped-wedge cluster randomized trials: a review and comparison [J].Int J Epidemiol,2022, 51(6):2000-2013.
- [24] Girling A, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models [J].Stat Med,2016, 35(13):66-2149.
- [25] Hooper R, Teerenstra S, deHoop E, et al. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials [J].Stat Med,2016, 35(26):4718-4728.
- [26] Li F, Turner E, Preisser J. Sample size determination for GEE analyses of stepped wedge cluster randomized trials [J].Biometrics,2018, 74(4):1450-1458.
- [27] Wang J, Cao J, Zhang S, et al. Sample size determination for stepped wedge cluster randomized trials in pragmatic settings [J].Stat Methods Med Res,2021, 30(7):1609-1623.
- [28] Kasza J, Hooper R, Copas A, et al. Sample size and power calculations for open cohort longitudinal cluster randomized trials [J].Stat Med,2020, 39(13):1871-1883.
- [29] Baio G, Copas A, Ambler G, et al. Sample size calculation for a stepped wedge trial [J].Trials,2015, 16:354.
- [30] Zhan D, Ouyang Y, Xu L, et al. Improving efficiency in the stepped-wedge trial design via Bayesian modeling with an informative prior for the time effects [J].Clin Trials, 2021, 18(3):295-302.
- [31] Xia F, Hughes J, Voldal E, et al. Power and sample size calculation for stepped-wedge designs with discrete outcomes [J].Trials, 2021, 22(1):598.
- [32] Kotwal S, Cass A, Coggan S, et al. Multifaceted intervention to reduce haemodialysis catheter related bloodstream infections: REDUCTION stepped wedge, cluster randomised trial [J].BMJ, 2022, 377: e069634.

(责任编辑:张悦)