心理学报 2026, Vol. 58, No. 4, 773−792 https://doi.org/10.3724/SP.J.1041.2026.0773
Acta Psychologica Sinica https://cstr.cn/32110.14.2026.0773

密集追踪干预研究设计中的建模及其样本量规划——基于动态结构方程模型*

刘 玥[1] 何月翎[1] 刘红云[2,3]

([1]四川师范大学脑与心理科学研究院, 成都 610066)
([2]应用实验心理北京市重点实验室; 3 北京师范大学心理学部, 北京 100875)

摘 要 密集追踪干预研究具有生态效度高、能够提供实时和个性化干预等优势。然而, 目前常用的数据分析方法未能充分反映密集追踪数据的特点, 而先进的数据分析模型又缺乏与之匹配的样本量规划方法, 极大地限制了这种范式的推广应用。本文在两种典型的密集追踪干预实验设计——单臂设计和随机对照设计下, 基于动态结构方程模型, 结合检验力和效应量估计准确性, 采用模拟研究方法开展样本量规划, 并从第一类错误率等方面对两种设计进行综合比较, 最后提出了实验设计和样本量规划建议。

关键词 密集追踪干预, 动态结构方程模型, 检验力分析, 效应量, 样本量规划

分类号 B841

1 引言

1.1 密集追踪干预研究的发展

随着心理学研究问题的不断深入, 传统基于实验室的干预研究面临生态效度低等挑战(Balaskas et al., 2021)。因此, 密集追踪干预(intensive longitudinal intervention)成为心理健康领域的热点范式(陈明瑞, 周萍, 2017; Schueller et al., 2017)。密集追踪干预, 又称生态瞬时干预(ecological momentary intervention, EMI), 是指在日常活动中提供实时的心理或行为障碍治疗措施, 帮助被试增强健康的心理状态和行为的频率(Heron & Smyth, 2010)。在干预内容上, 密集追踪干预既能对离散型行为(如成瘾行为)进行干预, 也能对连续型的日常情绪(如抑郁)、生理状态等进行干预。在干预的数据收集方案上, 密集追踪干预可以采用单独干预的形式只进行干预不开展密集追踪测量(如 Reininghaus et al., 2024), 也可以将密集追踪测量整合到干预中提供连续的实时监测和治疗支持(Bell et al., 2018)。

密集追踪干预起源于密集追踪研究 (intensive longitudinal study, ILS)。ILS 常采用生态瞬时评估(ecological momentary assessment, EMA)、经验取样(experience sampling method, ESM)或每日日记(daily diary, DD)等方法在自然状态下评估生活经历。这种方式能够对个体日常生活中的认知、情绪和行为等进行重复、实时的评估, 具有生态效度高、回忆偏差小的特点。因此, 近年来 ILS 的数量急剧增长(如 Wilhelm et al., 2012)。随后, 研究者把 ILS 的范式拓展到临床干预领域, 基于密集追踪测量收集的数据, 在日常生活中提供实时的、更具针对性的干预。Heron 和 Smyth (2010)首次明确提出生态瞬时干预(EMI)的概念。随着智能手机的普及, 移动应用程序成为了实施密集追踪干预最常见的媒介(Schueller et al., 2017)。在国外, 这种干预方式已经广泛应用于治疗焦虑、抑郁等心理障碍, 以及促进锻炼身体等健康行为以及吸烟等成瘾行为戒断(Smith & Juarascio, 2019), 均取得了良好效果。国内密集追踪干预研究虽处于起步阶段, 但是已呈现

收稿日期: 2025-02-26
* 国家自然科学基金项目(32200920,32471145), 四川省哲学社会科学基金青年人才项目(SCJJ25QN17)。
通信作者: 刘红云, E-mail: hyliu@bnu.edu.cn

出积极发展的态势(张祎 等, 2024)。此外, 近年来随着个性化干预需求的增加以及优化算法的快速发展, 密集追踪干预逐步演化为即时适应性干预(just-in-time adaptive interventions, JITAI; Schueller et al., 2017), 强调在自然情境下基于实时数据对干预内容进行动态调整。这一发展趋势要求在干预前后收集足够的密集追踪测量数据, 以多维度反映干预效果, 并为干预的持续优化与个体化反馈提供支持(Cuijpers et al., 2021)。总的来说, 尽管密集追踪干预研究的实践需求不断增多, 但国内外尚缺乏从方法学角度系统探讨如何通过样本量规划实现研究设计、以及如何基于数据分析对干预效果进行多维评价的研究, 极大限制了该领域研究的科学发展。

传统干预研究中两种典型的实验设计：单臂设计 (single-arm trial, SAT) 和随机对照设计 (randomized controlled trial, RCT), 可直接拓展于密集追踪干预研究(Yi, 2020)。SAT 是指所有被试都同时接受同样的干预处理, 其干预效果可以通过干预前阶段和干预阶段(简称前后测阶段)密集追踪测量结果的比较来评价。这种设计主要用于大型随机对照实验之前基于小样本的预研究(Baey & Le Deley, 2011), 或罕见病例等不易获取被试的研究(如老人, 见 Mair et al., 2022; 重度抑郁患者, 见 Rauschenberg et al., 2021; 问题赌博行为群体, 见 Hawker et al., 2021; 性风险行为群体, 见 Shrier & Spalding, 2017)。RCT 是指将所有被试随机分到干预组(实验组)或控制组, 干预组被试接受干预处理而控制组被试不接受任何处理或接受标准化处理(例如安慰剂), 其干预效果可以通过比较两组在前后测阶段密集追踪测量结果的差异来评价(例如 Bell et al., 2018)。这种设计具有两个明显优势：一是通过设置控制组能有效控制分离自然变化与外部干扰因素, 体现干预效应因果推断的严谨性; 二是通过随机分组降低选择偏差, 保证干预组和控制组在前测阶段的可比性, 更好满足因果推断的前提。

表 1 总结了 2015~2025 年间心理学领域 36 项密集追踪干预实证研究的基本情况(具体的检索说明见附录 https://doi.org/10.57760/sciencedb.psych.00506)。结果显示, 单臂设计和随机对照设计是两种主要的实验设计, 其中大部分研究(69.4%)采用了随机对照设计。

1.2 密集追踪干预研究的建模

已有密集追踪干预研究的数据分析主要采用了线性混合效应模型、方差分析、*t* 检验、回归分析等方法(见表 1)。但现有方法存在三个方面的明显不足。第一, 忽略了被试干预效应的个体差异。传统的方差分析、*t* 检验等方法没有考虑测量时间点嵌套于被试的层级结构, 忽略了干预效应可能存在的个体差异, 进而导致统计结论的偏差(Hoffman & Walters, 2022)。此外, 混合效应模型中的残差方差同质性假设在实际中很可能被违背, 个体内变异的异质性会对数据分析带来更大挑战(Hedeker et al., 2008)。第二, 忽略了时间序列数据的自相关结构。研究关心的结果变量(以下简称结果变量)被多次重复测量, 呈现出时间序列的特点, 而多数分析方法都未加入自相关, 会造成有偏差的参数估计结果, 引发错误的统计推论(Kenny & Judd, 1986)。第三, 仅关注从结果变量均值上反映出的干预效应。已有方法都以结果变量本身作为因变量建立回归模型, 未能充分结合变量动态发展的特点, 难以从多角度提供干预有效性的证据。

表 1 2015~2025 年心理学密集追踪干预研究总结(36 项)

对象	类型	数量(百分比)
实验设计	单臂设计	9(25.0%)
	随机对照设计	25(69.4%)
	微随机	2(5.6%)
统计建模	定性分析	1(2.8%)
	描述分析	2(5.5%)
	方差分析、*t* 检验、卡方检验, 回归分析	10(27.8%)
	主题分析法	1(2.8%)
	线性混合效应模型	20(55.6%)
	动态结构方程模型	2(5.5%)
样本量确定方式	检验力分析	11(30.6%)
	前人研究或预研究	6(16.6%)
	实际资源考虑	1(2.8%)
	未提及	18(50.0%)

注：微随机设计(microrandomized trials)是指被试在重复测量的试次中随机接受干预处理或不处理的实验设计。

因此, 为结合密集追踪干预研究的特点, 全面、多维评价干预效果, 有必要基于密集追踪常用的动态结构方程建模(dynamic structural equation modelling, DSEM)方法进行数据分析。DSEM 是一种集合了时间序列建模、结构方程建模和线性混合效应建模的分析方法(Asparouhov et al., 2018)。在密集追踪干预研究数据分析中, DSEM 的优势主要体现在两个方面：其一, 能充分反映数据本身的嵌套结构和时间序列特征; 其二, 可通过模型参数全面评价干预效果。这不仅包括结果变量均值的改变,

还包括其延续性(inertia)和个体内变异性(intra-individual variability, IIV)的改变(Sherwood, 2022)。例如, 负性情绪的延续性降低, 反映了个体从负性情绪中脱离的能力; 正性情绪的延续性增加, 反映了个体对正性情绪维持的能力, 可通过考察自相关系数前后测的变化来实现(Hamaker et al., 2021)。此外, 个体内变异性能够反映个体认知功能, 是很多身心障碍的重要预测因素(Aschenbrenner & Jackson, 2024)。个体变异性的改变可以通过考察个体内残差方差前后测的变化来实现。因此, 一些研究者拓展出了适用于密集追踪干预研究的 DSEM 模型。例如, Hamaker 等人(2021)在随机对照设计框架下, 拓展出从均值、延续性、个体内变异性三个方面评价干预效果的 DSEM 模型; Li 等人(2024)将自回归模型与潜变量增长曲线模型相结合, 探讨均值、自回归和个体内变异性在不同阶段的变化。但他们的方法对数据收集设计的要求较高, 需要大于两个阶段的密集追踪数据才能得到增长趋势相关的参数估计。此外, 线性增长(下降)的强假设不一定符合实际。

1.3　密集追踪干预研究的样本量规划

目前应用 DSEM 对密集追踪干预研究进行数据分析的研究屈指可数(见表 1, 仅 2 篇)。先进统计建模方法的应用, 以及结果的可重复性和推广性离不开与此相匹配的样本量规划方法的发展。随着学术界对心理学研究中学术不端和可重复性危机讨论日趋激烈, 越来越多的学术期刊和机构提倡合理应用样本量规划方法, 科学设置被试量、测量时间点数量等设计要素, 从而避免在不控制第一类错误率的情况下根据结果决定是否要继续收集数据直至达到显著的不良行为(例如 p-hacking), 促进科研过程和结果的公开透明, 创建更好的学术环境(胡传鹏 等, 2016; Nosek et al., 2022)。然而, 已有的密集追踪干预研究仅有少部分(30.6%)基于检验力分析确定样本量(见表 1), 一半(50.0%)的研究都未提及样本量确定依据, 这导致实际研究的被试量、干预次数(或与之匹配的密集测量次数)存在较大差异(见表 2)。

越来越多的研究提倡同时基于检验力和效应量准确性分析确定样本量(刘玥 等, 2024; Arend & Schäfer, 2019; Maxwell et al., 2008), 即科学、合理的样本量应同时保证: (1)正确拒绝虚无假设并估计效应的方向; (2)具有较高的效应量估计精度。一方面, 基于虚无假设显著性检验(Null Hypothesis Significance Test, NHST)的检验力分析要求样本量必须使检验力达到预设标准。在基于 DSEM 的检验力分析中, 由于其模型含多个随机参数, 采用公式推导的方法计算检验力非常困难, 现有研究多采用蒙特卡洛模拟的方式实现检验力分析进而确定样本量(Fang & Wang, 2024; Lafit et al., 2022; Schultzberg & Muthén, 2018), 但是很少有研究针对适用于密集追踪干预研究的 DSEM 模型开展检验力分析。另一方面, 效应量准确性分析的核心是控制效应量置信区间(Confidence Interval, CI)的宽度, 越窄表明其估计越准确(Maxwell et al., 2008)。此外, 参数估计值的相对偏差、置信区间对真值的覆盖率等指标也可以提供效应量估计的其他信息。因此, 刘玥等人(2024)提出了置信区间宽度等高线图, 帮助确定同时符合检验力、效应量准确性要求的样本量。不过目前尚没有研究基于针对适用于密集追踪干预研究的DSEM模型, 同时结合检验力和效应量准确性要求, 应用置信区间宽度等高线图提供样本量规划建议。

表 2　2015~2025 采用 SAT 和 RCT 设计的心理学密集追踪干预研究样本量情况(34 项)

样本量	实验设计	中数	最小值	最大值
每组被试量	SAT	31	10	152
	RCT	51	17	230
前测阶段测量次数(天数)	SAT	1(1)	1(1)	12(7)
	RCT	1(1)	1(1)	61(20)
干预阶段测量次数(天数)	SAT	84(21)	0(0)	128(84)
	RCT	60(14)	0(0)	392(192)
干预阶段干预次数(天数)	SAT	126(28)	24(7)	166(84)
	RCT	70(20)	14(7)	384(192)

注：表中仅总结了应用最广的两种设计(SAT 和 RCT)的样本量情况(34 项研究)。SAT 表示单臂设计, RCT 表示随机对照设计。由于随机对照设计一般采用两个随机对照组且每组被试量相等的设计, 因此表中统计每组被试量。有少数研究干预阶段没有开展密集追踪测量, 因此干预阶段测量次数(天数)的最小值为 0。

1.4　问题提出

综上, DSEM 模型是分析密集追踪干预研究数据的有效方法, 然而目前鲜有方法研究关注可应用于密集追踪干预数据分析的 DSEM 模型。Hamaker 等人(2021)所提出的 DSEM 模型不适用于单臂设计; Yi (2020)使用的模型仅关注了结果变量均值反映的干预效应。本研究基于密集追踪干预中最常见的单臂设计和随机对照设计, 对上述模型进行拓展, 使其能够全面的对均值、延续性、个体内变异性所体现的干预效应进行评价, 为个性化干预提供更详细的信息反馈。

目前, 很少有研究针对密集追踪干预研究实验设计, 探究基于 DSEM 的样本量规划问题。虽然 Yi (2020)在三种密集追踪干预实验设计下基于 DSEM 开展了样本量规划, 但该研究应用的模型仅关注结果变量均值所反映的干预效应, 而自回归、个体内变异性的变化对于评价干预效果同样重要。并且, 样本量规划的依据只参考了检验力分析, 没有结合效应量估计的结果(例如 Maxwell et al., 2008)。因此, 需要基于能够同时反映均值、自回归和个体内变异性干预效应的 DSEM 拓展模型, 探讨综合了检验力分析和效应量准确性分析的样本量规划问题, 为实际研究的样本量确定提供更科学的依据。

本文首先基于更科学的在干预前后均设置密集追踪测量的设计, 拓展并总结应用于单臂设计和随机对照设计下的 DSEM 模型。然后, 基于蒙特卡洛模拟方法, 探讨两种设计下 DSEM 模型的样本量规划, 并通过可信区间宽度等高线图直观呈现结合检验力分析和效应量准确性分析的样本量规划结果。其次, 采用模拟研究的方法, 从第一类错误率的角度在不同样本量条件下比较两种设计的差异。最后, 通过密集追踪干预的实证研究, 说明实际中如何基于 DSEM 模型实现样本量规划。本研究使用的所有 Mplus 代码见附录(https://doi.org/10.57760/sciencedb.psych.00506)。

2 应用于密集追踪干预研究的动态结构方程模型

在分析前, 将密集追踪干预研究收集的数据整理成适用于 Mplus 分析的长格式数据(见表 3)。其中 ID 表示被试编号(从 1 到 N),..表示被试在前测阶段测量的结果变量值, y_{post}表示被试在后测阶段测量的结果变量值, Time 表示测量时间点(前测从 1 到T_{pre}, 后测从 1 到T_{post}, $T = max\{T_{pre}, T_{post}\}$, 即, 在前后测测量时间点不同的情况下, 取最大的测量时间点 T, 未测量的时间点记为缺失), 前测阶段和后测阶段分别编码, Group 表示分组变量, 0 代表控制组, 1 代表干预组。单臂设计收集的数据包括第 1-4 列, 随机对照设计收集的数据包括第 1-5 列。

表 3 密集追踪干预研究数据示例

ID	y_{pre}	y_{post}	Time	Group
1	……	……	1	0
1	……	……	……	0
1	……	……	T	0
2	……	……	1	1
2	……	……	……	1
2	……	……	T	1
……	……	……	……	……
N	……	……	1	……
N	……	……	……	……
N	……	……	T	……

2.1 单臂设计下的动态结构方程建模

在 Yi (2020)提出的考察结果变量均值前后测差异模型的基础上, 本研究拓展了考察结果变量在干预前后测自回归和个体内变异性变化的 DSEM 模型(模型 1, 如图 1 所示)。

首先, 将结果变量 y 在前后测阶段的变异分别分解为个体间和个体内两部分。

分解模型:

前测阶段: $y_{pre.it} = \mu_{pre.i} + y^{W}_{pre.it}$, (1)

后测阶段: $y_{post.it} = \mu_{post.i} + y^{W}_{post.it}$, (2)

其中, $y_{pre.it}$表示被试 i 的结果变量在前测时间点 t 上的取值, $y_{post.it}$表示被试 i 的结果变量在后测时间点 t 上的取值(前后测阶段的时间点各自从 1 开始编码), $\mu_{pre.i}$和$\mu_{post.i}$表示个体间部分, $y^{W}_{pre.it}$和$y^{W}_{post.it}$表示个体内部分。然后, 分别定义前后测阶段的个体内模型。

个体内模型:

前测阶段 $y^{W}_{pre.it} = \varphi_{pre.i} y^{W}_{pre.it-1} + e_{pre.it}$, (3)

后测阶段$y^{W}_{post.it} = \varphi_{post.i} y^{W}_{post.it-1} + e_{post.it}$, (4)

其中, $\varphi_{pre.i}$和$\varphi_{post.i}$分别代表前测阶段和后测阶段随机的自回归系数; 个体内残差服从多元正态分布, 在实际应用中一般设定前后测残差相互独立, 即: $\begin{bmatrix} e_{pre.it} \\ e_{post.it} \end{bmatrix} \sim N\left(0, \Sigma_e = \begin{bmatrix} \sigma^2_{pre.i} & 0 \\ 0 & \sigma^2_{post.i} \end{bmatrix}\right)$。由于方差总是正值, 参考前人研究, 用个体内残差方差的对数形式(即$log(\sigma^2_{pre.i})$, $log(\sigma^2_{post.i})$)对个体内变异进行进一步建模(Hamaker et al., 2021; Schultzberg & Muthén, 2018)。个体间模型则分别构建了前测阶段模型和反映后测与前测差异的模型。

个体间模型:

前测阶段

$$\mu_{pre.i} = \gamma_{00} + u_{0i}, \quad (5)$$

$$\varphi_{pre.i} = \gamma_{10} + u_{1i}, \quad (6)$$

$$log(\sigma^2_{pre.i}) = \omega_{20} + u_{2i}, \quad (7)$$

后测与前测差异

$$\Delta\mu_i = \mu_{post.i} - \mu_{pre.i} = \gamma_{30} + u_{3i}, \quad (8)$$

$$\Delta\varphi_i = \varphi_{post.i} - \varphi_{pre.i} = \gamma_{40} + u_{4i}, \quad (9)$$

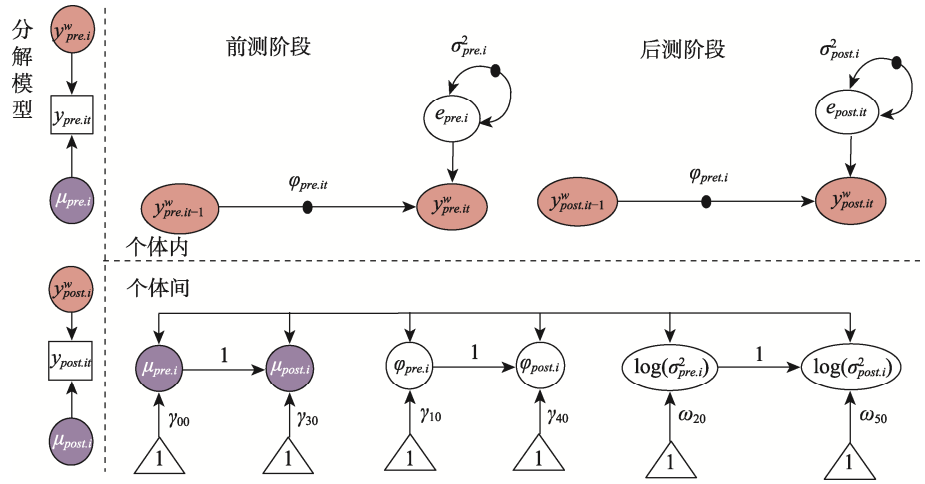


图 1　单臂设计的动态结构方程模型示意图

$$\Delta log(\sigma_i^2) = log(\sigma_{post.i}^2) - log(\sigma_{pre.i}^2) = \omega_{50} + u_{5i}, \quad (10)$$

其中，γ_{00}、γ_{10}和ω_{20}为前测阶段模型的固定参数，分别表示结果变量$y_{pre.it}$、自回归系数$\varphi_{pre.i}$和残差方差$log(\sigma_{pre.i}^2)$的个体间均值；γ_{30}，γ_{40}和ω_{50}为后测与前测阶段差异的固定参数，分别代表均值、自回归系数和个体内残差后测相对前测的变化，即三个方面的干预效应，也是检验力分析和效应量准确性分析针对的关键参数。值得注意的是，虽然理论上我们可同时定义(5)-(7)的个体间模型，但在参数估计时，只能估计前测阶段模型和前后测差异模型截距的固定效应、随机效应之和(例如公式(8)只能估计$\gamma_{00} + \gamma_{30}$，$u_{0i} + u_{3i}$)，而无法分离前测效应和干预效应。因此，在实际应用中可以采用两步法实现模型的参数估计。第一步，先用贝叶斯方法对前测模型中的参数进行估计，并保存其后验分布；第二步，从第一步得到的参数估计值的后验分布中抽取若干个参数取值(例如 100 个)，将其视为固定值，估计前后测差异模型中的差异参数，从而获得差异参数的后验分布。在基于蒙特卡洛模拟的检验力分析中，可将前测部分的固定效应(γ_{00}，γ_{10}和ω_{20})设定为 0，以便检验反映干预效应的固定效应(γ_{30}，γ_{40}和ω_{50})。个体间模型的随机参数服从多元正态分布

$$\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \\ u_{3i} \\ u_{4i} \\ u_{5i} \end{bmatrix} \sim N\left(0, \Sigma_u = \begin{bmatrix} \tau_{00}^2 & & & & & \\ \tau_{01} & \tau_{11}^2 & & & & \\ \tau_{02} & \tau_{12} & \tau_{22}^2 & & & \\ \tau_{03} & \tau_{13} & \tau_{23} & \tau_{33}^2 & & \\ \tau_{04} & \tau_{14} & \tau_{24} & \tau_{34} & \tau_{44}^2 & \\ \tau_{05} & \tau_{15} & \tau_{25} & \tau_{35} & \tau_{45} & \tau_{55}^2 \end{bmatrix}\right)。$$

值得说明的是，单臂设计假设结果变量不存在前后阶段自然的变化，前后阶段的所有变化均归因于干预效应，实际研究中这个假设可能过于严苛。另外，个体间的模型也有另一种等价形式，可以将前后测差异的方程(8)-(10)重构为对后测阶段建模(Mplus 估计采用的模型)。例如，公式(8)变为$\mu_{post.i} = \gamma_{30} + \gamma_{00} + u_{3i}^{'}$。可知$u_{3i}^{'} = u_{0i} + u_{3i}$，当所有随机效应独立时，后测结果的随机效应包含前测的随机效应(u_{0i})和差异部分的随机效应(u_{3i})。

2.2　随机对照设计下的动态结构方程模型建模

随机对照设计下的动态结构方程模型(模型 2)采用了 Hamaker 等人(2021)拓展的形式，如图 2 所示。

在后测与前测差异的个体间模型，即方程(8)-(10)中加入分组变量 Group (如 Group = 1 表示干预组，Group = 0 表示控制组)，通过分组变量对后测与前测差异的均值、自回归和个体内变异性(即$\Delta\mu_i$，$\Delta\varphi_i$，$\Delta log(\sigma_i^2)$)预测的回归系数反映干预效果。模型 2 的分解模型和个体内模型与单臂设计下相同，即与方程(1)-(4)相同。对于个体间模型，前测阶段模型也与(5)-(7)相同，不同的是随机对照设计下后测与前测差异模型增加了分组变量，即：

$$\Delta\mu_i = \mu_{post.i} - \mu_{pre.i} = \gamma_{03} Group_i + u_{3i}, \quad (11)$$

$$\Delta\varphi_i = \varphi_{post.i} - \varphi_{pre.i} = \gamma_{41} Group_i + u_{4i}, \quad (12)$$

$$\Delta log(\sigma_i^2) = log(\sigma_{post.i}^2) - log(\sigma_{pre.i}^2) = \omega_{05} Group_i + u_{5i}, \quad (13)$$

其中，$Group_i$是虚拟变量，代表每名被试 i 属于干预组还是控制组，γ_{03}，γ_{41}和ω_{05}分别代表分组变量对均值、自回归和个体内残差前后测差异的影响，即干预效应，也是检验力分析和效应量准确性分析针对的关键参数。与单臂设计下的模型相同，个体间模型的随机参数服从多元正态分布。值得注意的是，该模型假设对于控制组，结果变量不存在前后测阶段的差异。并且与模型 1 类似，当所有随机效应部分独立时，后测结果的随机效应包含前测的随机效应和差异部分的随机效应。

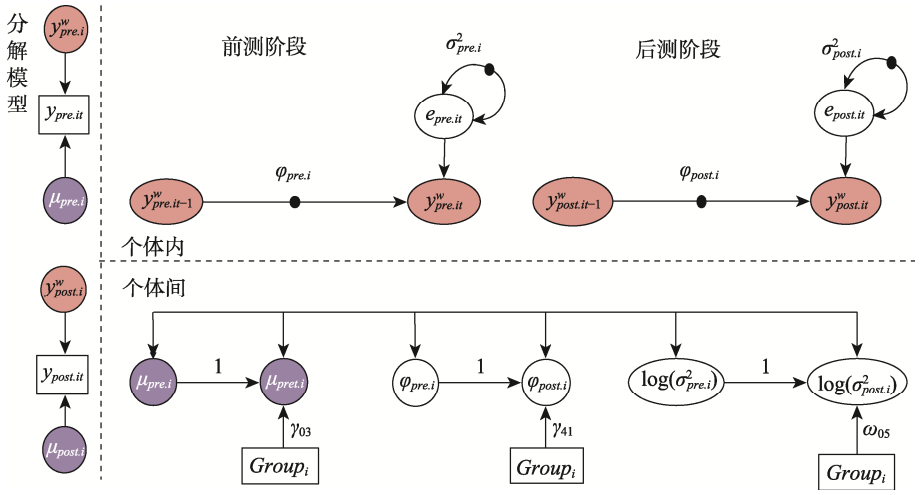


图 2 随机对照设计的动态结构方程模型示意图

实际中，随着时间推移，结果变量在前后测阶段可能发生自然的变化(Hamaker et al., 2021)。如果采用单臂设计，这种自然变化会混淆在实验处理效应中无法分离。而采用随机对照设计可以通过控制组前后测的差异分离出自然变化，再基于分组变量的回归系数得到更加纯粹的干预效应。本文进一步将随机对照设计下的DSEM模型，拓展存在前后测自然变化的模型(模型 3)。与上面模型的区别是，后测与前测模型表述如下：

$$\Delta\mu_i = \mu_{post.i} - \mu_{pre.i} = \mu_{change.i} + \gamma_{03}Group_i + u_{3i}, \quad (14)$$

$$\Delta\varphi_i = \varphi_{post.i} - \varphi_{pre.i} = \varphi_{change.i} + \gamma_{41}Group_i + u_{4i}, \quad (15)$$

$$\Delta log(\sigma_i^2) = log(\sigma_{post.i}^2) - log(\sigma_{pre.i}^2) = log(\sigma_{change.i}^2) + \omega_{05}Group_i + u_{5i}, \quad (16)$$

其中，带有下标 $change.i$ 的项表示控制组的个体 i 在干预前后阶段的差异，即两个阶段相比的自然变化。其余参数的含义与模型 2 相同。

3 模拟研究 1：单臂设计下的样本量规划

模拟研究 1 基于 SAT，在 2.1 建构的模型框架下，针对γ_{30}，γ_{40}和ω_{50}三个反映干预效果的参数，在不同样本量条件下考察其检验力和参数估计准确性，并通过可信区间宽度等高线图提供样本量建议。

3.1 模拟研究设计

本研究考虑了两种情境。情境 1：前后测阶段测量点数量相等；情境 2：前后测阶段测量点数量不等。实际研究中前测阶段测量时间点数一般少于后测阶段(例如 Mair et al., 2022)。但是，考虑到前测阶段也需要保证足够数量的测量时间点以得到准确、稳定的自回归、个体内变异性参数估计结果。因此，本研究以前后测阶段测量时间点数为 1:3 的情况为例，考察测量时间点数量是否为平衡设计对样本量规划的影响。

3.1.1 样本量水平

参考密集追踪干预实证研究设计(表 1)以及DSEM 相关模拟研究设计设置样本量水平(Fang & Wang, 2024; Schultzberg & Muthén, 2018)。总的测量时间点数(T)包含 6 个水平：20, 40, 80, 120, 160 和 200。被试量(N)包含 7 个水平：30, 60, 100,150, 200, 300 和 400。二者完全交叉共形成$6 \times 7 = 42$种样本量组合。

3.1.2 数据生成

参考前人类似研究的参数设置(Fang & Wang, 2024; Schultzberg & Muthén, 2018; Yi, 2020)，基于模型 1 生成数据。设定固定效应$\gamma_{00} = 0$, $\gamma_{10} = 0.2$, $\omega_{20} = 0$。本研究出于简化考虑，限定所有随机部分独立，即假设随机效应 $\begin{bmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \\ u_{3i} \\ u_{4i} \\ u_{5i} \end{bmatrix} \sim N\left(0, \Sigma_u = \begin{bmatrix} \tau_{00}^2 & & & & & \\ 0 & \tau_{11}^2 & & & & \\ 0 & 0 & \tau_{22}^2 & & & \\ 0 & 0 & 0 & \tau_{33}^2 & & \\ 0 & 0 & 0 & 0 & \tau_{44}^2 & \\ 0 & 0 & 0 & 0 & 0 & \tau_{55}^2 \end{bmatrix}\right)$。前测阶段的随机效应参考以往研究设置为$\tau_{00}^2 = \tau_{11}^2 = \tau_{22}^2 = 0.04$(Fang & Wang, 2024; Yi, 2020)。参考密集追踪

干预实证研究(见第 6 部分)所得到的两个部分随机效应估计值的比较, 并结合不同参数设置的预研究, 将前后测差异模型的随机效应设置为$\tau_{33}^2 = \tau_{44}^2 = \tau_{55}^2 = 0.01$。参考样本量规划的类似研究(Arend & Schäfer, 2019), 干预效应量γ_{30}, γ_{40}和ω_{50}的值代表了前后测的差异, 根据 Cohen (1988)的标准设定为中等水平, 即标准化的参数值为$\gamma_{std} = 0.5$ (参考 Cohen's d = 0.2, 0.5, 0.8)。由于本研究仅关注所有个体平均的干预效应, 因此基于个体间水平的随机效应方差对其标准化(Arend & Schäfer, 2019), 得到产生数据的参数值$\gamma_{generate} = \gamma_{std} \times \sqrt{var}$, 其中 var 是个体间水平总的随机效应方差。基于本研究所有随机效应独立的设定, 则 var 为个体间水平所有随机效应的方差之和, 因此, $\gamma_{30} = \gamma_{40} = \omega_{50} = \gamma_{generate} = 0.5 \times \sqrt{0.04 + 0.01} = 0.112$。每种样本量组合条件下, 数据重复模拟生成 500 次。

3.1.3　拟合模型

采用与产生数据模型相同的模型(模型 1), 基于 DSEM 参数估计常采用的贝叶斯估计方法拟合模型并估计参数。注意模拟研究中假设每个密集追踪测量阶段内的数据符合平稳性假设, 因此不需要进行去趋势处理(Zhou et al., 2021)。但实际研究仍需要对每个阶段采取去趋势的处理(见第 6 部分)。数据模拟和分析均应用 Mplus8.10 完成(Muthén & Muthén, 1998-2017)。具体来说, 使用 Mplus 中的 MONTECARLO 命令定义产生数据的情况(例如样本量、重复次数等)。使用 MODEL POPULATION 命令定义产生数据的模型及参数真值。使用 ANALYSIS 命令定义拟合数据的模型。在贝叶斯估计中, 采用 Mplus 默认的无信息先验(即回归系数的固定部分$\gamma \sim N(0, +\infty)$, 随机效应的方差$\sigma^2 \sim \Gamma^{-1}(-1,0)$)。设置最小迭代次数为 10000, 默认前 5000 次为 burn-in, 链条数采用默认值 2。收敛标准采用 Mplus 的默认设置, 即每个参数的 PSR (potential scale reduction)值小于 1.1 则认为收敛。

3.1.4　评价标准

评价指标包括 4 个方面。(1)收敛率。即参数估计收敛次数占总重复次数的比例。后面的所有评价指标均基于收敛的情况计算。(2)检验力。反映干预效应的参数(γ_{30}, γ_{40}和ω_{50})的 95%可信区间(以下简称可信区间)不包括 0 的次数占所有重复次数的比例。预设的检验力标准为大于等于 0.8。(3)效应量估计准确性[1]。包括反映感兴趣的干预效应的参数的相对估计偏差(relative parameter estimation bias, rbias), 误差均方根(root mean squared error, RMSE), 可信区间宽度(width), 可信区间对真值的覆盖率(CP)。(4)效应量标准误估计准确性。即, 反映感兴趣的干预效应的参数的估计标准误相对于其估计值标准差的偏差(SE-SD bias)。如果效应量估计准确, rbias 应在[−0.1, 0.1]的区间内, RMSE 应较小, width 应较窄, CP 应在 0.925 到 0.975 之间(Bradley, 1978)。如果估计标准误准确, SE-SD bias 应接近 0 (Schultzberg & Muthén, 2018)。

3.2　研究结果

在平衡设计情境下, 在所有条件下模型收敛率都达到 95%以上, 其中有 81%的条件下收敛率为 100%。在非平衡设计下, N = 30, T = 20 的条件下收敛率仅为 49%, 其余条件下收敛率都达到 90%以上, 其中有 48%的条件下收敛率为 100% (结果见网络版附表 2)。平衡设计下的收敛率优于非平衡设计。

3.2.1　检验力

平衡设计情况下各参数在不同样本量条件下的检验力结果如表 4 所示。从表中可以看出, 检验力随着样本量增加而增加。当被试量达到一定水平后(如$N \geq 200$), 测量时间点数量对检验力的影响变小。三个参数检验力水平基本一致, 相同样本量条件下, 与自回归干预效应相关的参数γ_{40}的检验力大于另外两个参数, 并且在 N 小时差异最大; 而与个体内变异性干预效应相关的参数ω_{50}的检验力水平最低。非平衡设计情况下的检验力略低于平衡设计(网络版附表 3)。

3.2.2　效应量及其标准误估计准确性

表 4 呈现了平衡设计情况下不同样本量条件下三个参数的相对偏差、可信区间宽度结果, 其余结果(误差均方根、覆盖率和标准误偏差)见网络版附表 4。从表中可以看出, 所有条件下γ_{30}的相对偏差都在可接受范围内, 而当测量时间点数量较少时(例如 T = 20 或者 40), 其余两个参数相对偏差会小于−0.1, 低估干预效应量。可信区间宽度随样本量

[1] 一般来说, 准确性(accuracy)指估计值与真值的接近程度, 包括 rbias, RMSE 等指标, 精确性(precision)指估计值的稳定性, 主要指标准误 standard error (SE), 反映为置信区间宽度。但是在模型设定正确(如本研究中参数估计的模型与产生数据模型相同), 参数估计方法无偏的情况下, 二者近似等价。因此, 我们参考 Maxwell 等(2008)提出的概念 accuracy in parameter estimation (AIPE), 统称为准确性。

表 4 单臂设计平衡设计情况下检验力、相对偏差、可信区间宽度结果

N	T	power			rbias			width		
		γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}
30	20	**0.088**	**0.118**	**0.080**	−0.044	**−0.202**	**−0.256**	0.534	0.397	0.547
	40	**0.178**	**0.250**	**0.180**	0.024	**−0.143**	−0.086	0.385	0.270	0.379
	80	**0.358**	**0.608**	**0.336**	−0.024	−0.019	−0.048	0.271	0.191	0.270
	120	**0.474**	**0.766**	**0.480**	−0.001	−0.039	−0.039	0.225	0.158	0.222
	160	**0.624**	0.854	**0.564**	−0.013	−0.029	−0.037	0.197	0.141	0.196
	200	**0.704**	0.936	**0.658**	−0.012	−0.006	−0.029	0.181	0.131	0.178
60	20	**0.200**	**0.262**	**0.126**	−0.027	**−0.163**	**−0.231**	0.358	0.270	0.371
	40	**0.400**	**0.552**	**0.328**	0.026	**−0.115**	−0.080	0.255	0.183	0.257
	80	**0.668**	0.928	**0.666**	−0.032	−0.002	−0.017	0.180	0.130	0.183
	120	0.850	0.986	0.836	0.012	−0.018	−0.024	0.149	0.107	0.150
	160	0.922	0.996	0.912	−0.016	−0.020	−0.024	0.131	0.095	0.132
	200	0.966	1.000	0.966	−0.010	−0.010	−0.010	0.120	0.089	0.120
100	20	**0.344**	**0.474**	**0.202**	−0.017	**−0.143**	**−0.186**	0.271	0.205	0.283
	40	**0.660**	0.848	**0.538**	0.018	−0.077	−0.074	0.193	0.139	0.196
	80	0.888	0.996	0.882	−0.018	−0.014	0.009	0.135	0.098	0.139
	120	0.966	1.000	0.964	−0.013	−0.010	−0.004	0.113	0.082	0.114
	160	0.996	1.000	0.996	−0.002	−0.011	−0.013	0.099	0.073	0.100
	200	1.000	1.000	0.998	−0.013	−0.005	−0.010	0.091	0.067	0.091
150	20	**0.560**	**0.652**	**0.364**	0.011	**−0.117**	**−0.155**	0.219	0.166	0.228
	40	0.806	0.960	**0.750**	0.006	−0.063	−0.042	0.154	0.112	0.158
	80	0.972	0.992	0.974	−0.009	−0.024	0.007	0.109	0.079	0.112
	120	0.998	0.998	0.996	0.002	−0.020	−0.001	0.090	0.067	0.092
	160	1.000	1.000	1.000	−0.014	−0.020	0.004	0.080	0.060	0.081
	200	1.000	1.000	1.000	0.007	−0.012	0.001	0.073	0.055	0.074
200	20	**0.652**	0.808	**0.496**	0.001	−0.096	**−0.125**	0.188	0.142	0.196
	40	0.898	0.986	0.872	−0.010	−0.048	−0.033	0.133	0.097	0.137
	80	0.996	1.000	0.994	−0.012	−0.008	−0.007	0.094	0.069	0.096
	120	0.998	1.000	1.000	−0.006	−0.005	−0.002	0.078	0.057	0.079
	160	1.000	1.000	1.000	−0.011	−0.002	−0.010	0.069	0.051	0.070
	200	1.000	1.000	1.000	−0.011	0.000	−0.003	0.063	0.047	0.063
300	20	0.812	0.940	**0.690**	−0.023	−0.066	**−0.103**	0.153	0.116	0.160
	40	0.974	0.998	0.972	−0.013	−0.037	−0.039	0.108	0.079	0.111
	80	1.000	1.000	1.000	−0.015	0.000	0.004	0.076	0.056	0.078
	120	1.000	1.000	1.000	−0.010	−0.002	0.004	0.063	0.047	0.064
	160	1.000	1.000	1.000	−0.011	0.002	−0.010	0.056	0.042	0.057
	200	1.000	1.000	1.000	−0.006	−0.002	−0.005	0.051	0.038	0.052
400	20	0.926	0.982	0.846	−0.013	−0.051	−0.079	0.131	0.100	0.137
	40	0.998	0.998	0.992	−0.012	−0.028	−0.036	0.092	0.068	0.096
	80	1.000	1.000	1.000	−0.010	0.001	0.006	0.066	0.048	0.067
	120	1.000	1.000	1.000	−0.006	0.000	0.001	0.055	0.040	0.056
	160	1.000	1.000	1.000	−0.010	−0.001	−0.009	0.048	0.036	0.049
	200	1.000	1.000	1.000	−0.003	−0.001	−0.002	0.044	0.033	0.045

注：N 表示被试量, T 表示测量时间点总数量, power 表示检验力, 其中加粗的表示小于 0.8 的结果, rbias 表示相对偏差, 其中加粗的表示在[−0.1,0.1]范围外的结果, width 表示可信区间宽度。

增加而减小。误差均方根的结果与相对偏差较为一致。覆盖率基本都介于 0.925 和 0.975 之间, 说明覆盖率较好, 标准误适中。标准误偏差也集中在 0 附近波动, 说明效应量标准误估计较准确。非平衡设计情况下的效应量及其标准误估计结果见网络版附表 3 和附表 5。非平衡设计下的相对偏差略小于平衡设计, γ_{30}和γ_{40}的相对偏差整体较小, ω_{50}的相对偏差会在 $T=20$ 的情况下超出可接受范围。在被试量较小的条件下(N = 30, 60)非平衡设计的可信区间宽度略大于平衡设计。非平衡设计下的误差均方根、覆盖率和标准误偏差与平衡设计相当。

3.2.3　样本量规划建议

本研究参考刘玥等(2024)提出的置信区间宽度等高线图给出样本量规划建议。由于本研究采用贝叶斯估计得到参数的后验分布可信区间, 因此将置信区间宽度等高线图改称为可信区间宽度等高线图。图 3 和网络版附图 1、2 分别表示平衡设计下γ_{30}, γ_{40}和ω_{50}三个参数的可信区间宽度等高线图(代码见网络版附录)。网络版附图 3、4、5 表示非平衡设计下三个参数的可信区间宽度等高线图。其中, 阴影区域表示符合检验力大于等于 0.8 标准的条件, 不同颜色的等高线对应于不同的可信区间宽度。参考刘玥等(2024)的方法, 根据效应量 Cohen's d (1988)小和大的标准值 0.2 和 0.8, 定义可接受的最宽可信区间宽度。效应量小的条件下$\gamma_{generate} = \gamma_{std} \times \sqrt{var} = 0.2 \times \sqrt{0.04 + 0.01} = 0.045$, 效应量大的条件下 $\gamma_{generate} = \gamma_{std} \times \sqrt{var} = 0.8 \times \sqrt{0.04 + 0.01} = 0.179$, 因此最宽可信区间宽度为 0.179−0.045 = 0.134。

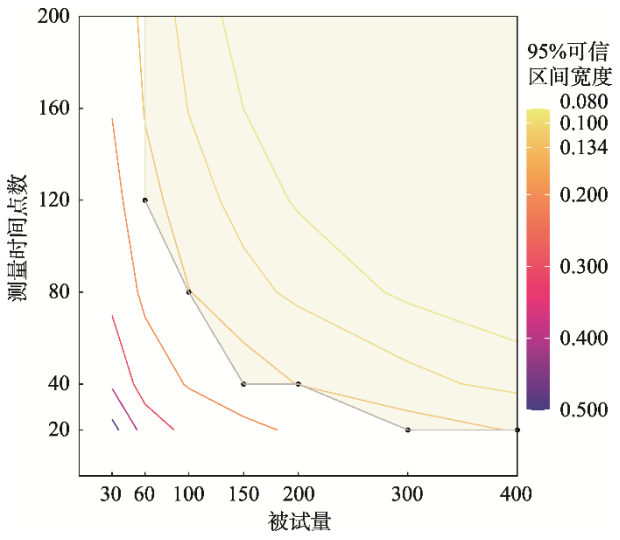


图 3　单臂设计平衡设计下γ_{30}的可信区间宽度等高线图
注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。如图例所示从 0.08 到 0.50, 在图中共有 7 条依次排列的等高线。例如, 0.134 对应的等高线表示线条以上的区域 95% CI 宽度在 0.134 及其以下。彩图见电子版, 下同。

在平衡设计下, 从图中可以看出, 被试量和测量时间点数具有相互补偿作用, 增加任一者的数量都会增加检验力, 减少可信区间宽度。但是, 被试量需满足一定要求(如对于γ_{30}至少为 60), 才能使得检验力可能大于 0.8。阴影区域的边缘和最宽可信区间宽度(0.134)等高线接近重合, 说明根据检验力和效应量估计准确性的要求确定的样本量具有高度一致性。对比三个参数的可信区间宽度等高线图可以发现, γ_{40}的阴影区域最大, 另外两个参数较为接近, 说明γ_{40}对样本量的需求相对较小。根据可信区间宽度等高线图, 可以在阴影区域边缘的转折处结合可信区间宽度等高线, 并考虑具体研究成本, 得到合适的样本量。

具体来看, 在效应量中等的条件下, 如果同时采用检验力大于等于 0.8 和可信区间宽度小于 0.134 的标准, 对于γ_{30}, 推荐的被试量为 100, 测量时间点数为 80 (8000 个数据点); 或被试量为 60, 测量时间点数为 160 (9600 个数据点)。对于γ_{40}, 推荐的被试量为 60, 测量时间点数为 80 (4800 个数据点)。对于ω_{50}, 推荐的被试量为 60, 测量时间点数为 160 (9600 个数据点)。综合来看, 如果需要正确识别三个参数所反映的干预效应并且准确估计效应量, 应取三个参数所需样本量的最大值, 平衡设计下至少需要 60 名被试和 160 个测量时间点; 非平衡设计下至少需要 150 名被试和 80 个测量时间点, 相比于平衡设计需要更大的 N 和更小的 T。

4　模拟研究 2：随机对照设计密集追踪干预研究样本量规划

模拟研究 2 基于 RCT, 在 2.2 建构的模型框架下, 针对反映干预效果的三个参数γ_{03}, γ_{41}和ω_{05}, 在不同样本量条件下考察其检验力和参数估计准确性, 并通过可信区间宽度等高线图提供样本量建议。与研究 1 相同, 本研究也考虑了前后测阶段测量点数量相等(情境 1, 1 : 1)和前后测阶段测量点数量不等(情境 2, 1 : 3)两种情境。

4.1　模拟研究设计

模拟研究 2 的样本量水平与模拟研究 1 相同, 共包含 42 种样本量组合。采用干预组与控制组平衡(被试量)相等的平衡设计(每组被试量为 N/2), 因此样本量中的被试量水平包括两组被试的总人数。

基于模型 2 生成模拟数据, 其中分组变量$Group_i$为包含两个类别的分类变量。由于本研究采用模型 2, 即以分组变量为水平 2 自变量的模型, 因此参考 Rights 和 Sterba (2019)在线性混合效应模型(多水平模型)中对每一层模型分别构建效应量R^2的思路, 对于每种干预效应的水平 2 回归模型, 计算分组变量固定斜率所解释的变异占该层变异的比例R^2, 并将其设为中等水平 0.09 (Cohen, 1988)。以均值干预效应为例(公式(11)), 将模型变换为后测均值为因变量的形式, 即$\mu_{post.i} = \mu_{pre.i} + \gamma_{03}Group_i + u_{3i}$。本研究假设所有随机效应独立, 则该回归方程对因变量的解释程度$R^2 = \gamma_{03}^2 var(Group_i)/(\tau_{00}^2 + \tau_{33}^2 + \gamma_{03}^2 var(Group_i))$。由于本研究采用两组被试量相等的设计, 因此分组变量的方差$var(Group_i) = p \times (1-p) = 0.5 \times 0.5 = 0.25$, p 代表每组的人数占总人数的比例。参考模拟研究 1 的设置, 所有随机部分限定为独立, $\tau_{00}^2 = 0.04$, $\tau_{33}^2 = 0.01$, 可得到$\gamma_{03} = 0.140$。同理可以求出γ_{41}和ω_{05}的值均为 0.140[2]。生成数据的其余参数真值与研究相同。拟合数据的模型与产生数据模型相同。Mplus 中贝叶斯估计的设置、评价标准等均与模拟研究 1 相同。

4.2 研究结果

在所有条件下模型收敛率都达到 95%以上, 其中有 88%的条件下收敛率为 100%。

4.2.1 检验力

平衡设计下各参数在不同样本量条件下的检验力结果如表 5 所示。与模拟研究 1 结果类似, 检验力随着样本量增加而增加。当被试量达到一定水平后(如$N \geq 200$), 测量时间点数量对检验力的影响变小。相同样本量条件下, 与个体内变异性干预效应相关的参数ω_{05}的检验力小于另外两个参数。总体而言, 模拟研究 2 得到的检验力都小于模拟研究 1。这可能是因为在 RCT 下需要把所有被试平均分成两组, 减少了每组的样本量。如果考虑单组被试量相等的条件(例如模拟研究 2 中 $N = 60$ 和模拟研究 1 中 $N = 30$ 的条件), 模拟研究 2 得到的检验力与模拟研究 1 水平大致相当。在 T 小的条件下($T = 20$), 非平衡设计情况下的检验力大于平衡设计, 这种现象在 N 大($N \geq 150$)且 $T = 20$ 的条件下更明显(网络版附表 6)。

4.2.2 效应量及其标准误估计准确性

表 5 呈现了平衡设计下不同样本量条件下三个参数的相对偏差、可信区间宽度结果, 其余结果(误差均方根、覆盖率和标准误偏差)见网络版附表 7。从表中可以看出, 除了在被试量为 30 和 60, 测量时间点数为 20 的条件下会低估反映均值干预效应的参数γ_{03}, 其余条件下各参数的相对偏差都在可接受范围内。可信区间宽度随样本量增加而减小。误差均方根的结果与相对偏差较为一致。覆盖率基本都介于 0.925 和 0.975 之间, 在测量时间点数偏少的条件下(例如$T \leq 80$), 覆盖率高于 0.975 的情况稍多, 说明在这些情况下参数估计的标准误可能较大。标准误偏差集中在 0 附近波动, 说明效应量标准误估计较准确。非平衡设计情况下的效应量及其标准误估计结果见网络版附表 6 和附表 8。非平衡设计下的可信区间宽度略小于平衡设计, 这种现象在 N 和 T 小的条件下更加明显。非平衡设计下的相对偏差、误差均方根、覆盖率和标准误偏差与平衡设计相当。

4.2.3 样本量规划建议

首先参照模拟研究 1 的方法计算三个参数的最宽可信区间宽度。参考刘玥等(2024)的方法, 根据效应量R^2小和大的标准值 0.01 和 0.25, 定义可接受的最宽可信区间宽度。根据 4.1 中的方法求出效应量小的条件下回归系数的值为 0.045, 效应量大的条件下回归系数的值为 0.258, 因此最宽可信区间宽度为 $0.258-0.045 = 0.213$。图 4 和网络版附图 6、7 分别表示平衡设计下γ_{03}, γ_{41}和ω_{05}三个参数的可信区间等高线图。网络版附图 8、9、10 分别表示非平衡设计下三个参数的可信区间等高线图。

在平衡设计下, 从图中可以看出, 与模拟研究 1 相似, 在被试量达到一定数量后, 被试量和测量时间点数具有相互补偿作用。例如, 对于γ_{03}和ω_{05}, 要达到检验力要求必须至少有 100 名被试; 对于γ_{41}, 必须至少有 80 名被试; 这高于模拟研究 1 中需要的最小被试量。此外, 反映自回归干预效应的参数阴影区域最大, 对样本量的需求相对较小。与模拟研究 1 不同, 模拟研究 2 可信区间等高线图中的阴影区域的面积减小, 说明符合检验力要求需要更大被试量。模拟研究 2 中最宽可信区间宽度的等高线都在阴影区域下边缘以下, 说明这一指标比检验力要求的样本量小。

[2] 在本研究模拟设置分组变量只有两个类别的情况下, 也可以采用与研究 1 类似的方法, 设定差异类效应量(Cohen's d, 即考察干预组和控制组在前后测差异上的差异)为中等水平来确定产生模型的真值。得到反映干预效应的三个参数(γ_{03}, γ_{41} 和 ω_{05})的真值为 0.112, 与 0.140 非常接近。

表 5　随机对照设计平衡设计下检验力、相对偏差、可信区间宽度结果

N	T	power			rbias			width		
		γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}
30	20	**0.038**	**0.070**	**0.048**	**−0.134**	0.017	0.004	0.933	0.676	0.932
	40	**0.080**	**0.172**	**0.108**	−0.013	0.014	0.004	0.679	0.482	0.662
	80	**0.150**	**0.302**	**0.164**	−0.035	0.003	−0.019	0.498	0.364	0.492
	120	**0.242**	**0.366**	**0.216**	0.009	−0.040	−0.003	0.425	0.309	0.421
	160	**0.316**	**0.492**	**0.278**	0.042	0.005	−0.047	0.374	0.281	0.373
	200	**0.326**	**0.550**	**0.344**	0.044	−0.016	0.020	0.345	0.261	0.345
60	20	**0.100**	**0.212**	**0.136**	**−0.103**	0.054	0.053	0.598	0.441	0.598
	40	**0.218**	**0.432**	**0.232**	0.016	0.030	−0.010	0.435	0.319	0.433
	80	**0.350**	**0.648**	**0.376**	−0.047	0.007	−0.013	0.324	0.241	0.323
	120	**0.511**	**0.781**	**0.513**	−0.025	0.015	0.033	0.273	0.207	0.276
	160	**0.625**	0.826	**0.582**	0.026	−0.003	−0.040	0.246	0.187	0.247
	200	**0.706**	0.890	**0.700**	0.005	−0.003	0.008	0.226	0.174	0.227
100	20	**0.218**	**0.398**	**0.224**	−0.037	0.054	0.024	0.444	0.329	0.445
	40	**0.380**	**0.664**	**0.382**	0.005	0.029	0.001	0.325	0.240	0.323
	80	**0.594**	0.878	**0.624**	−0.044	0.007	−0.002	0.242	0.181	0.243
	120	**0.760**	0.940	**0.782**	−0.029	0.005	0.019	0.205	0.156	0.208
	160	0.859	0.975	0.832	0.002	0.002	−0.019	0.184	0.141	0.186
	200	0.921	0.992	0.902	−0.005	−0.009	0.013	0.170	0.131	0.171
150	20	**0.308**	**0.584**	**0.344**	−0.057	0.029	0.033	0.355	0.264	0.356
	40	**0.516**	0.850	**0.562**	−0.023	0.036	0.024	0.261	0.192	0.260
	80	**0.788**	0.974	**0.798**	−0.042	−0.001	0.004	0.194	0.146	0.195
	120	0.902	0.988	0.926	0.007	−0.021	0.026	0.165	0.126	0.167
	160	0.962	0.996	0.956	−0.002	0.011	−0.018	0.147	0.115	0.149
	200	0.986	1.000	0.988	0.019	−0.005	0.020	0.136	0.107	0.138
200	20	**0.376**	**0.692**	**0.444**	−0.042	0.000	0.050	0.304	0.226	0.305
	40	**0.710**	0.926	**0.726**	−0.012	0.031	0.029	0.223	0.165	0.224
	80	0.912	0.992	0.936	−0.032	0.002	0.017	0.166	0.125	0.168
	120	0.978	1.000	0.968	0.006	−0.016	0.023	0.141	0.109	0.144
	160	0.990	1.000	0.990	−0.001	0.006	−0.003	0.126	0.099	0.129
	200	0.998	1.000	0.998	0.010	−0.003	0.022	0.117	0.092	0.119
300	20	**0.560**	0.874	**0.636**	−0.053	0.030	0.031	0.245	0.182	0.246
	40	0.860	0.994	0.884	−0.031	0.011	0.008	0.180	0.134	0.180
	80	0.982	1.000	0.984	0.004	0.012	−0.006	0.135	0.102	0.135
	120	0.998	1.000	0.994	0.004	−0.012	0.015	0.114	0.088	0.117
	160	1.000	1.000	1.000	−0.001	−0.004	−0.001	0.102	0.080	0.104
	200	1.000	1.000	1.000	0.000	−0.001	0.003	0.095	0.075	0.096
400	20	**0.688**	0.942	**0.778**	−0.049	0.040	0.038	0.211	0.156	0.212
	40	0.944	0.996	0.962	−0.020	0.026	0.034	0.155	0.115	0.156
	80	0.994	1.000	1.000	−0.013	−0.002	0.018	0.116	0.088	0.117
	120	1.000	1.000	1.000	0.000	−0.008	0.013	0.099	0.076	0.101
	160	1.000	1.000	1.000	0.000	−0.003	0.001	0.088	0.069	0.090
	200	1.000	1.000	1.000	0.005	−0.003	0.018	0.082	0.065	0.083

注：N 表示被试量(N/2 表示每组被试量), T 表示测量时间点总数量, power 表示检验力, 其中加粗的表示小于 0.8 的结果, rbias 表示相对偏差, 其中加粗的表示在[−0.1,0.1]范围外的结果, width 表示可信区间宽度。

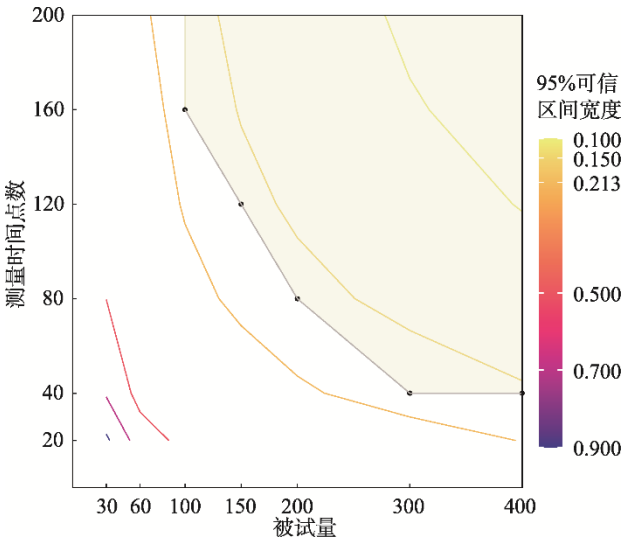


图 4 随机对照设计平衡设计下γ_{03}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。

具体来看，在效应量中等的条件下，如果同时采用检验力大于等于 0.8 和可信区间宽度小于 0.213 的标准，推荐采用两组被试量相等的平衡设计，对于γ_{03}，总被试量为 100，测量时间点数为 160 (16000 个数据点)；或总被试量为 150，测量时间点数为 120 (18000 个数据点)。对于γ_{41}，推荐的总被试量为 60，测量时间点数为 160 (9600 个数据点)；或者总被试量为 100，测量时间点数为 80 (8000 个时间点)。对于ω_{05}，推荐的总被试量为 100，测量时间点数为 160 (16000 个数据点)；或总被试量为 150，测量时间点数为 120(18000 个数据点)。综合来看，如果需要正确识别三个参数所反映的干预效应并且准确估计效应量，平衡设计下至少需要 100 名被试和 160 个测量时间点；非平衡设计下至少需要 100 名被试和 120 个测量时间点。非平衡设计需要的测量时间点更少。

5 模拟研究 3：两种设计下的第一类错误率比较

结合两个模拟研究的结果发现，为达到检验力和效应量准确性要求，采用 SAT 所需要的样本量小于采用 RCT 需要的样本量。然而，这并不意味着在实际研究中推荐采用 SAT。由于 SAT 无法分离自然变化与干预效应，在实际干预研究中可能会导致干预效应的错误估计。模拟研究 3 考虑包含前后测自然变化的情境，在前后测阶段测量点数量相等的情境下，分别采用两种设计下的 DSEM 模型分析数据，比较不同样本量条件下干预效应参数的第一类错误率，以进一步说明两种密集追踪干预研究设计的适用条件。

5.1 模拟研究设计

5.1.1 数据生成

基于模型 3 生成模拟数据。分组变量对干预阶段均值、自回归和个体内变异性的干预效应γ_{03}，γ_{41}和ω_{05}设定为 0，代表不存在干预效应。随机效应部分采用与模型 2 相同的设置。代表自然变化的参数μ_{change}，φ_{change}和$log(\sigma^2_{change})$设定为两个水平：一是根据 Cohen (1988)的标准设定为效应量小的水平($\gamma_{std}=0.2$)，即$\gamma_{generate}=\gamma_{std}\times\sqrt{var}=0.2\times\sqrt{0.04+0.01}=0.045$；二是比前者更小的水平($\gamma_{std}=0.1$)，即$\gamma_{generate}=\gamma_{std}\times\sqrt{var}=0.1\times\sqrt{0.04+0.01}=0.022$。

5.1.2 模拟因素

结合本研究目的，考虑三个模拟因素。自然变化的效应量(ES)包含 2 个水平：0.1 和 0.2。由于本研究并不关注样本量规划问题，因此不需要像模拟研究 1、2 那样覆盖尽可能多的样本量水平。因此，研究三采用相对前两个模拟研究较少的样本量水平，前后测阶段总的测量时间点数(T)包含 5 个水平：40, 80, 120, 160 和 200，被试量(N)包含 3 个水平：60，100 和 200。三个因素完全交叉共形成 2 × 5 × 3 = 30 种样本量组合。每种样本量组合条件下，数据重复模拟生成 500 次。

5.1.3 拟合模型

对于 RCT，采用模型 3 拟合产生数据。对于 SAT，在模拟产生的数据中选取干预组被试，应用模型 1 拟合数据。因此，SAT 下的被试量仅为 RCT 的一半。Mplus 中贝叶斯估计的设置与模拟研究 1、2 相同。

5.1.4 评价指标

主要参考第一类错误率对两种设计下的分析结果进行比较，即在所有收敛的重复中，反映干预效应的参数(SAT 的γ_{30}，γ_{40}和ω_{50}，RCT 的γ_{03}，γ_{41}和ω_{05})的可信区间不包括 0 的次数占所有收敛次数的比例。理想的第一类错误率应在[0.025，0.075]范围内。

5.2 研究结果

两种设计下相应模型分析的第一类错误率结果如表 6 所示。从表中可以看出，当存在前后测阶段自然变化时，采用 RCT 得到的第一类错误率基本均在可接受范围，但是采用 SAT 的第一类错误率在被试量和测量时间点数较大的情况下均超过 0.05，存在第一类错误率高估的情况。并且，这种

表 6　两种设计下第一类错误率结果

N	*T*	0.1						0.2					
		mean		autoregression		IIV		mean		autoregression		IIV	
		SAT	RCT	SAT	RCT	SAT	RCT	SAT	RCT	SAT	RCT	SAT	RCT
60(30)	40	**0.024**	**0.022**	0.030	**0.022**	0.032	0.034	0.044	**0.022**	0.050	**0.022**	0.050	0.034
60(30)	80	0.046	0.048	0.030	0.024	0.036	0.036	**0.088**	0.040	**0.102**	0.032	0.054	0.030
60(30)	120	0.048	0.044	0.068	0.048	0.052	0.036	**0.088**	0.040	**0.170**	0.038	**0.096**	0.032
60(30)	160	0.034	0.028	0.066	0.036	0.046	0.046	**0.154**	0.044	**0.196**	0.036	**0.120**	0.042
60(30)	200	**0.076**	0.040	**0.086**	0.068	0.072	0.058	**0.152**	0.050	**0.216**	0.040	**0.148**	0.038
100(50)	40	0.056	0.034	0.034	0.036	0.044	0.048	**0.102**	0.032	**0.094**	0.038	**0.060**	0.046
100(50)	80	0.070	0.056	0.062	0.036	0.038	0.042	**0.136**	0.036	**0.190**	0.050	**0.106**	0.034
100(50)	120	0.056	0.060	**0.090**	0.050	**0.078**	0.034	**0.184**	0.042	**0.290**	0.056	**0.166**	0.034
100(50)	160	0.070	0.042	**0.132**	0.058	0.074	0.058	**0.244**	0.036	**0.306**	0.038	**0.214**	0.040
100(50)	200	**0.104**	0.046	**0.118**	0.052	0.068	0.038	**0.284**	0.054	**0.392**	0.048	**0.246**	0.044
200(100)	40	**0.088**	0.044	**0.080**	0.044	0.042	0.064	**0.180**	0.044	**0.208**	0.038	**0.092**	0.052
200(100)	80	**0.086**	0.048	**0.106**	0.044	0.072	0.048	**0.294**	0.062	**0.400**	0.048	**0.202**	0.040
200(100)	120	**0.114**	0.064	**0.178**	0.046	**0.092**	0.060	**0.380**	0.050	**0.518**	0.048	**0.326**	0.054
200(100)	160	**0.146**	0.048	**0.218**	0.046	**0.134**	0.042	**0.486**	0.044	**0.624**	**0.076**	**0.404**	0.044
200(100)	200	**0.178**	0.048	**0.234**	0.070	**0.146**	0.050	**0.522**	0.040	**0.710**	0.040	**0.504**	0.038

注：0.1 和 0.2 分别表示前后测自然变化的效应量, mean 表示均值干预效应, autoregression 表示回归的干预效应, .IIV 表示个体内变异性, SAT 表示单臂设计, RCT 表示随机对照设计, *N* 表示被试量, 其中括号外的数值表示随机对照设计下(两组)的被试量, 括号内的数值表示单臂设计的被试量, *T* 表示测量时间点总数量, 加粗的表示第一类错误率在[0.025, 0.075]范围之外的结果。

高估随着前后测自然变化效应量的增加而增加。三种干预效应中, 自回归的干预效应第一类错误率高估最为严重, 与这个系数的检验力相对最高的结果一致(模拟研究 1、2)。具体来看, 当自然变化的效应量为 0.2 时, SAT 在被试量为 30 的情况下就出现了第一类错误率高估的情况。当被试量为 100, 测量时间也为 200 的情况下, 三个参数第一类错误率都达到0.5以上。当自然变化的效应量为0.1时, SAT 在被试量为 50 的情况下出现了第一类错误率高估的情况。当被试量为 100, 测量时间也为 200 的情况下, 三个参数第一类错误率都达到 0.14 以上。

6　实证研究：基于社交媒体外貌焦虑生态瞬时干预的样本量规划

实证研究基于一项社交媒体外貌焦虑生态瞬时干预的预实验, 说明在实际中如何运用本研究使用的方法指导样本量规划。具体包括以下步骤：(1) 开展预实验并分析结果, 确定模拟研究生成数据的真值; (2) 应用蒙特卡洛模拟方法基于 DSEM 和设定的真值, 重复生成数据并分析; (3) 计算包括检验力和效应量准确性在内的评价指标; (4) 综合各评价指标结果, 绘制可信区间宽度等高线图, 确定合适的样本量。

6.1　预实验及其结果分析

研究发现, 一些基于视觉内容的高度视觉化社交媒体的使用会引发一种短暂的外貌状态焦虑, 称为社交媒体外貌焦虑(Hawes et al., 2020)。我们基于随机对照设计, 对女大学生的社交媒体外貌焦虑采用基于自我同情语句阅读的方式开展了生态瞬时干预, 并得到了初步研究结果。后续研究计划基于更广泛的包含不同性别的样本, 对预实验结果进行进一步验证。因此需要通过科学的样本量规划对扩大样本群体的正式研究设计合理的样本量。预实验流程及结果见附录。

6.2　基于预实验的样本量规划

以预实验结果(网络版附表 11)作为参数真值, 参考模拟研究 2 生成数据、拟合模型并分析。注意基于预实验的样本量规划可以在参考预实验结果的基础上进行尝试, 在模拟研究中设置合适的样本量水平。由于模拟研究发现被试量为 100 的条件下, 即使 T = 200, 所有反映干预效应参数的检验力都小于 0.8。因此被试量(*N*)设置 4 个水平：200, 250, 300 和 400。前后测阶段总的测量时间点数(*T*)设置 5 个水平：40, 80, 120, 160 和 200。三个反映干预效应的参数的检验力和效应量准确性结果见网络版附表 9、10。从表中可以看出, 当测量时间点数

量较小时($T \leq 40$)，γ_{41}的相对偏差小于−0.1，倾向于低估该参数。其他参数的相对偏差都在[−0.1,0.1]的范围内。RMSE 和相对偏差结果较为一致。覆盖率的值都在[0.925, 0.975]范围内，说明覆盖率较好，标准误适中。标准误偏差集中在 0 附近波动，说明效应量标准误估计较准确。由于γ_{41}真值的效应量很小，即使样本量增加到 N = 400，T = 200 也无法使其检验力大于 0.8。因此，无法呈现该参数的可信区间宽度等高线图。γ_{03}和ω_{05}的可信区间宽度等高线图见图 5 和图 6。从图中可以看出，对于均值干预效应，当被试量为 250，测量时间点数为 80 (20000 个数据点)时，或被试量为 300，测量时间点数为 40 (12000 个数据点)时，可以保证检验力大于 0.8。对于个体内变异性的干预效应，当被试量为 200，测量时间点数为 120 (24000 个数据点)时，或当被试量为 250，测量时间点数为 40 (10000 个数据点)时，可以保证检验力大于 0.8。最后，研究者可以根据所期望达到的参数估计可信区间宽度，在阴影范围内结合可信区间等高线选择合适的样本量。

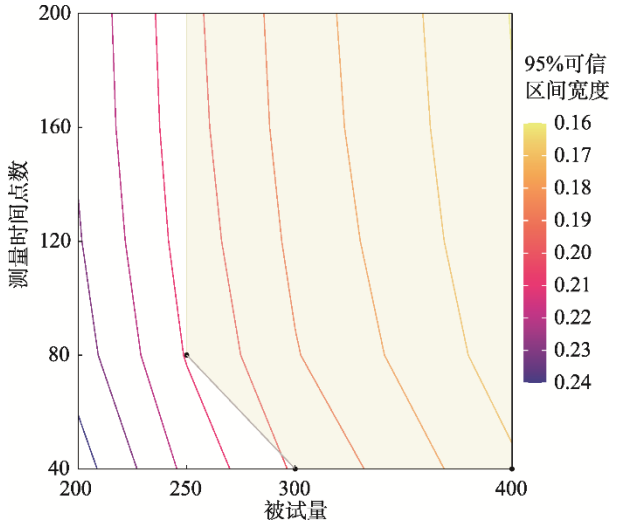


图 5 基于预实验结果的随机对照设计下γ_{03}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。

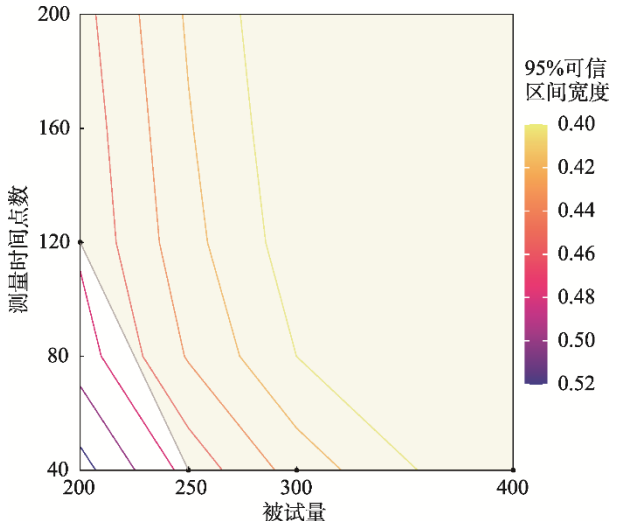


图 6 基于预实验结果的随机对照设计下ω_{05}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。

7 讨论

7.1 样本量规划主要评价指标结果

研究发现，应用 DSEM 模型分析数据基本都能成功收敛。在单臂设计下，满足可信区间宽度要求样本量稍大于满足检验力要求的样本量；在随机对照设计下，满足可信区间宽度要求样本量小于满足检验力要求的样本量。说明在不同情况下，基于检验力分析和效应量准确性分析确定的样本量可能不同。研究者可以根据自己的具体需要，选用一种或两种标准确定样本量。此外，基于本研究得到的推荐样本量发现实证研究中部分样本量普遍偏小(见表 2)，这一方面说明部分实证研究没有科学论证样本量规划依据，另一方面可能是由于实证研究中的数据分析模型更简单，所需样本量也更小。从样本量规划的评价指标来看，两种设计下要达到理想的检验力，都至少需要被试量满足一定要求。参数估计准确性普遍较好。对于单臂设计，在测量时间点数量较少的情况下(例如 T = 20 或 40)会低估反映自回归和个体内变异性干预效应的效应量。对于随机对照设计，在被试量和测量时间点数都较少(例如 N = 30 或 60，T = 20)的情况下会低估反映均值干预效应的效应量。参数估计值可信区间对真值的覆盖率较好，参数估计标准误适中。参数标准误估计较准确。注意均值、自回归、残差方差的量尺不同，导致在同等效应量条件下得到的检验力和参数估计准确性结果不一定可比。例如，在实际中基于 DSEM 的平稳性假设，可能很难得到对于自回归系数中等甚至大的干预效应。因此对不同干预效应样本量规划结果的比较应特别谨慎。在是否为前后测平衡设计的比较方面，对于单臂设计，非平衡设计的检验力略低于平衡设计，可信区间宽度在被试量较小的条件下略大于平衡设计。对于随机对照设计，非平衡设计在测量时间点数少(T = 20)的情况下检验力大于平衡设计，可信区间宽度略小于平衡设计。总的来说，在测量时间点数少的情况下，非平衡设计更具优势，这样可以在有限的 T 中将更多

的测量用于更重要的干预阶段。

7.2　两种设计下样本量规划的比较

本研究探索了两种设计下的样本量规划问题, 并比较了两种设计下的第一类错误率结果。研究发现, 在干预效应量相等的条件下, 单臂设计对样本量的要求小于随机对照设计。但是, 如果被试在前后测阶段存在结果变量随时间的自然变化, 则采用单臂设计无法控制这种差异, 会高估干预效应的第一类错误率, 这种现象在样本量大的情况下更严重。

其实两种设计有各自的优缺点及适用情境。单臂设计所需样本量虽然较小, 但后测密集追踪测量阶段被试在心理障碍症状上的改善有可能是自身发展变化的结果, 而非单纯的干预效应, 具有较高第一类错误率的风险。随机对照设计更加科学, 通过加入控制组分离随着时间推进其他因素对干预效果的影响, 得到更纯粹的干预效应, 这在持续时间较长的密集追踪干预研究中显得尤为重要(Hamaker et al., 2021)。因此这种设计在密集追踪干预研究中应用最广(74.3%, 见表 1)。但是, 随机对照设计对样本量, 尤其是被试量的要求高于单臂设计, 极大的提高了研究成本, 降低密集追踪干预研究的可行性。

7.3　实践建议

本研究在两种典型的密集追踪干预设计下探讨了样本量规划问题, 并对不同样本量条件下的第一类错误率进行了比较。应用研究者可以参考本研究附录中提供的Mplus代码生成不同样本量组合条件下的模拟数据、分析数据并计算相关的评价指标, 基于这些结果再调用刘玥等(2024)研究中提供的语句方便地绘制可信区间宽度等高线图(见附录), 获得推荐样本量。基于研究结果提出以下建议供实践应用者参考。

第一, 结合具体情境选择合适的干预实验设计。通过对不同设计下样本量规划的比较可知, 单臂设计和随机对照设计具有各自的优缺点和适用情境。研究者应根据具体的结果变量、被试人群、测量频率等特征, 选择合适的干预实验设计。当干预的被试人群较易获得, 或者干预研究需要设置多个(两个及以上)对照组时, 建议选择更加科学的随机对照设计(Wright et. al., 2018)。当干预的被试人群或者症状较稀少, 或已知结果变量不会发生前后测的自然变化时, 可以采用样本量要求较低的单臂设计。但是, 无论采用何种设计, 都应该基于科学的方法充分论证样本量规划的理由。

第二, 根据密集追踪数据特点选择科学的数据分析方法。尽管目前动态结构方程模型已经日益成为密集追踪研究的主流数据分析方法, 但是还没有将这一方法应用到密集追踪干预领域的实证研究, 仅有少量研究从方法学角度验证了 DSEM 模型用于考察干预效应的有效性(Hamaker et al., 2021; Yi, 2020)。为充分挖掘密集追踪干预研究中获得的密集追踪数据, 准确反映数据的时间序列、嵌套结构特性, 从均值、自回归、个体内变异性三个角度多维评价干预效果, 我们建议采用本研究使用的DSEM 模型分析数据。同时, 也应该基于这些模型开展样本量规划, 通过科学的样本量规划助力先进数据分析方法的推广应用。需要注意的是, 本研究提出的样本量规划方法仅关注了与干预效应相关的参数, 如果在实际中还关注其他参数的估计结果, 或者其他问题(如模型拟合、模型预测准确性等), 则需要综合考虑相应的评价指标开展样本量规划。

第三, 基于模拟的方法开展样本量规划需注意产生模型的参数设置。应用模拟方法进行检验力和效应量准确性分析需要设定产生数据模型的各参数值。由于干预效应是个体层面变量, 只有固定效应可以参考样本量规划的类似研究(Arend & Schäfer, 2019)将其设定为中等水平。但是, 三个反映干预效应的因变量(均值、自回归和残差方差)的可能取值范围不同, 并且采用不同的效应量指标所得到的小、中、大效应的标准会有所区别, 给固定效应的参数设置带来困难。此外, DSEM 模型包含多项随机效应, 其产生值的设置仅有少量研究可供参考。并且, 经过预研究发现, 随机效应方差的设置会直接影响模型中干预效应参数估计的准确性。因此, 设置合适的随机效应方差也是大部分基于DSEM 的模拟研究重点关注的问题(例如 Fang & Wang, 2024)。综上, 我们建议参考前人开展的涉及结果变量的密集追踪研究所得到的固定效应和随机效应, 设置产生模型的各参数值。在没有参考研究的情况下, 可以开展预实验, 结合估计得到的固定效应和随机效应的结果设置产生值。但是, 也有研究指出基于预实验有偏的点估计值无法得到合理的样本量(例如 Albers & Lakens, 2018)。因此, 可以考虑从预实验估计效应量的分布中抽取真值, 结合具体研究对最小效应量的要求(smallest effect size of interest, SESOI), 参考元分析结果, 或直接应用本研究参数设置的方式来设定产生模型的参数值。但是无论采用何种方法, 都需要在正式研究

中明确说明模型产生值设置的依据，以规范样本量规划流程。

第四，在样本量规划中全面考虑时间趋势问题。密集追踪干预研究中可能包含两种时间趋势效应，一是在干预前后的每个阶段内，结果变量可能随时间变化。这种趋势违背了稳定性假设(Zhou et al., 2021)，会导致模型参数估计的偏差。在密集追踪研究的模拟研究中，一般假设数据符合平稳性假设而不做去趋势处理(例如 Asparouhov et al., 2018; Fang & Wang, 2024)。因此，本研究在模拟研究部分都没有采用去趋势化的模型。而在基于实证研究的样本量规划中，一般会在实证数据分析的部分采用去趋势处理(例如 Wang & Maxwell, 2015)。例如，实证研究中前后测阶段得到了显著为负的时间效应($\gamma_{前测时间效应} = -0.012$，CI = [−0.016, −0.008]; $\gamma_{前测时间效应} = -0.007$, CI = [−0.011, −0.003])，说明社交媒体外貌焦虑随时间有下降的趋势。二是干预前后两个阶段之间的自然变化，可以理解为在不采取任何干预措施的情况下(控制组)，结果变量的均值、自回归系数和个体内变异性在前后测阶段出现的差异。对于这种效应，我们建议在实验设计上分离其对干预效应的干扰。综上，我们建议条件允许的情况下尽可能采用随机对照设计，并且在应用模拟的方法开展样本量规划时，不需要去趋势。但在分析实际数据时，需要对各阶段内进行去趋势。

最后，我们建议在实际密集追踪干预研究中采用以下流程图的顺序确定合理的样本量(见图 7)。如果各干预效应考虑中等程度的效应量，并且参考本研究其他参数的真值设置，要同时满足检验力和效应量估计准确性要求，我们建议单臂设计至少采用 $N = 60$, $T = 160$ (平衡设计)或 $N = 150$, $T = 80$ (前后测 1:3 非平衡设计)，随机对照设计至少采用 $N = 100$, $T = 160$ (平衡设计)；或 $N = 100$, $T = 120$ (前后测 1 : 3 非平衡设计)。

7.4 未来研究展望

本研究具有一定的局限性，因此未来研究可以从三个方面开展。第一，本研究参考前人做法(刘玥等, 2024; Usami, 2020)根据期望的可信区间上下限，倒推可接受的最大可信区间宽度。但是当效应量为

图 7 密集追踪干预研究样本量规划流程图

小或者大水平(例如 Cohen's d = 0.2 或者 0.8), 或者不在临界值上时(例如 Cohen's d = 0.4 或者 0.6), 就无法根据这种方法确定最宽的可信区间宽度。近期, Kowialiewski (2025)的研究提出了基于效应量稳定性标准来确定样本量大小的方法, 即通过模拟的方式不断增加样本量直至效应量的参数估计值的变化小于预设的阈值。未来研究可以参考这种方式确定使得效应量估计值稳定的最宽可信区间宽度, 以帮助规划样本量。第二, 本研究采用了频率统计样本量规划的思想, 基于效应量的固定取值开展蒙特卡洛模拟研究, 忽略了效应量的不确定性(例如 Pek & Park, 2019)。未来研究可以在贝叶斯方法的框架下, 基于特定分布生成效应量, 得到更加科学的推荐样本量分布(Kruschke & Liddell, 2018)。第三, 随着模型复杂程度增加, 已有基于模拟的样本量规划方法的计算量呈几何级增长, 分析效率低。并且, 设置离散水平的样本量只能得到有限条件的组合, 可能无法得到全局最优解。为解决这一问题, 有研究者引入了基于机器学习预测的代理模型框架进行检验力分析, 并通过搜索算法得到最合理的样本量设计方案(Zimmer & Debelak, 2025)。未来研究可以将这种方法应用到基于 DSEM 的密集追踪干预研究样本量规划中, 并开发相应的应用软件, 促进研究的高效、广泛开展。

参　考　文　献

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187−195.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, *24*(1), 1−19.

Aschenbrenner, A. J., & Jackson, J. J. (2024). High-frequency assessment of mood, personality, and cognition in healthy younger, healthy older and adults with cognitive impairment. *Aging, Neuropsychology, and Cognition*, *31*(5), 914−931.

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(3), 359−388.

Baey, C., & Le Deley, M. C. (2011). Effect of a misspecification of response rates on type I and type II errors, in a phase II Simon design. *European Journal of Cancer*, *47*(11), 1647−1652.

Balaskas, A., Schueller, S. M., Cox, A. L., & Doherty, G. (2021). Ecological momentary interventions for mental health: A scoping review. *PloS One*, *16*(3), e0248152.

Bell, I. H., Fielding-Smith, S. F., Hayward, M., Rossell, S. L., Lim, M. H., Farhall, J., & Thomas, N. (2018). Smartphone-based ecological momentary assessment and intervention in a coping-focused intervention for hearing voices (SAVVy): Study protocol for a pilot randomised controlled trial. *Trials*, *19*(1), 262.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144−152.

Chen, M., & Zhou, P. (2017). Ecological momentary assessment and intervention of substance use. *Advances in Psychological Science*, *25*(2), 247−252.

[陈明瑞, 周萍. (2017). 成瘾物质使用的生态瞬时评估与干预. *心理科学进展*, *25*(2), 247−252.]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Erlbaum.

Cuijpers, P., Pineda, B. S., Quero, S., Karyotaki, E., Struijs, S. Y., Figueroa, C. A., ... Muñoz, R. F. (2021). Psychological interventions to prevent the onset of depressive disorders: A meta-analysis of randomized controlled trials. *Clinical Psychology Review*, *83*, 101955.

Fang, Y., & Wang, L. (2024). Dynamic structural equation models with missing data: Data requirements on N and T. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(5), 891−908.

Hamaker, E. L., Asparouhov, T., & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., p. 31). Guilford Press.

Hawes, T., Zimmer-Gembeck, M. J., & Campbell, S. M. (2020). Unique associations of social media use and online appearance preoccupation with depression, anxiety, and appearance rejection sensitivity. *Body Image, 33*, 66−76.

Hawker, C. O., Merkouris, S. S., Youssef, G. J., & Dowling, N. A. (2021). A smartphone-delivered ecological momentary intervention for problem gambling (GamblingLess: Curb Your Urge): Single-arm acceptability and feasibility trial. *Journal of Medical Internet Research*, *23*(3), e25786.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed‐effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627−634.

Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, *15*(1), 1−39.

Hoffman, L., & Walters, R. W. (2022). Catching Up on Multilevel Modeling. *Annual Review of Psychology*, *73*, 659−689.

Hu, C., Wang, F., Guo, J., Song, M., Sui, J., & Peng, K. (2016). The replication crisis in psychological research. *Advances in Psychological Science*, *24*(9), 1504−1518.

[胡传鹏, 王非, 过继成思, 宋梦迪, 隋洁, 彭凯平. (2016). 心理学研究中的可重复性问题: 从危机到契机. *心理科学进展, 24*(9), 1504−1518.]

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, *99*(3), 422−431.

Kowialiewski, B. (2025). The power of effect size stabilization. *Behavior Research Methods*, *57*(1), 7.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178−206.

Lafit, G., Sels, L., Adolf, J. K., Loeys, T., & Ceulemans, E. (2022). PowerLAPIM: An application to conduct power analysis for linear and quadratic longitudinal actor-partner interdependence models in intensive longitudinal dyadic

designs. *Journal of social and Personal Relationships*, *39*(10), 3085−3115.

Li, Y., Williams, L., Muth, C., Heshmati, S., Chow, S. M., & Oravecz, Z. (2024). A growth of hierarchical autoregression model for capturing individual differences in changes of dynamic characteristics of psychological processes. *Structural Equation Modeling: A Multidisciplinary Journal*, *32*(2), 237−250.

Liu, Y., Xu, L., Liu, H., Han, Y., You, X., & Wan, Z. (2024). Confidence interval width contours: Sample size planning for linear mixed-effects models. *Acta Psychologica Sinica*, *56*(1), 124−138.

[刘玥, 徐雷, 刘红云, 韩雨婷, 游晓锋, 万志林. (2024). 置信区间宽度等高线图在线性混合效应模型样本量规划中的应用. *心理学报*, *56*(1), 124−138.]

Mair, J. L., Hayes, L. D., Campbell, A. K., Buchan, D. S., Easton, C., & Sculthorpe, N. (2022). A personalized smartphone-delivered just-in-time adaptive intervention (JitaBug) to increase physical activity in older adults: Mixed methods feasibility study. *JMIR Formative Research*, *6*(4), e34662.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, *59*, 537−563.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide*. Muthén and Muthén.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*, 719−748.

Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*, *24*(5), 590−605.

Rauschenberg, C., Boecking, B., Paetzold, I., Schruers, K., Schick, A., van Amelsvoort, T., & Reininghaus, U. (2021). A compassion-focused ecological momentary intervention for enhancing resilience in help-seeking youth: uncontrolled pilot study. *JMIR Mental Health*, *8*(8), e25650.

Reininghaus, U., Daemen, M., Postma, M. R., Schick, A., Hoes-van der Meulen, I., Volbragt, N., ... van Amelsvoort, T. (2024). Transdiagnostic ecological momentary intervention for improving self-esteem in youth exposed to childhood adversity: The SELFIE randomized clinical trial. *JAMA Psychiatry*, *81*(3), 227−239.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309−338.

Schueller, S. M., Aguilera, A., & Mohr, D. C. (2017). Ecological momentary interventions for depression and anxiety. *Depression and Anxiety*, *34*(6), 540−545.

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 495−515.

Sherwood, S. N. (2022). *Feasibility and efficacy of virtual darkness in reducing intra-individual sleep variability among young adults with insomnia* [Unpublished Doctoral dissertation]. University of Nevada, Las Vegas.

Shrier, L. A., & Spalding, A. (2017). “Just take a moment and breathe and think”: Young women with depression talk about the development of an ecological momentary intervention to reduce their sexual risk. *Journal of Pediatric and Adolescent Gynecology*, *30*(1), 116−122.

Smith, K. E., & Juarascio, A. (2019). From ecological momentary assessment (EMA) to ecological momentary intervention (EMI): Past and future directions for ambulatory assessment and interventions in eating disorders. *Current Psychiatry Reports*, *21*(7), 53.

Usami, S. (2020). Confidence interval‐based sample size determination formulas and some mathematical properties for hierarchical data. *British Journal of Mathematical and Statistical Psychology*, *73*(S1), 1−31.

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*(1), 63−83.

Wilhelm, P., Perrez, M., & Pawlik, K. (2012). Conducting research in daily life: A historical review. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 62−86). New York: Guilford Press.

Wright, C., Dietze, P. M., Agius, P. A., Kuntsche, E., Livingston, M., Black, O. C., … Lim, M. S. (2018). Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: A three-armed randomized controlled trial. *JMIR Mhealth Uhealth, 6*(7), e149.

Yi, Z. (2020). *Intensive longitudinal data analyses and sample size considerations in intervention studies with dynamic structural equation modeling* [Unpublished Doctoral dissertation]. University of South Florida.

Zhang, W., Xu, L., Pan, X., Yao, L., Zhong, W., & Li, J. (2024). Application progress of ecological momentary intervention in health behavior promotion. *Journal of Nursing Science*, *39*(2), 116−121.

[张祎, 徐岚, 潘习, 姚林, 仲伟莹, 李佳璇. (2024). 生态瞬时干预在健康行为促进中的应用进展. *护理学杂志*, *39*(2), 116−121.]

Zhou, L., Wang, M., & Zhang, Z. (2021). Intensive longitudinal data analyses with dynamic structural equation modeling. *Organizational Research Methods*, *24*(2), 219−250.

Zimmer, F., & Debelak, R. (2025). Simulation-based design optimization for statistical power: Utilizing machine learning. *Psychological Methods, 30*(3), 513−536.

Data analysis and sample size planning for intensive longitudinal intervention studies using dynamic structural equation modeling

LIU Yue[1], HE Yueling[1], LIU Hongyun[2,3]

([1] *Institute of Brain and Psychological Sciences, Sichuan Normal University, Chengdu* 610066, *China*)

([2] *Beijing Key Laboratory of Applied Experimental Psychology, Beijing Normal University, Beijing* 100875, *China*)

([3] *Faculty of Psychology, Beijing Normal University, Beijing* 100875, *China*)

Abstract

Intensive longitudinal interventions (ILIs) have emerged as powerful tools for understanding, treating and preventing mental and behavioral disorders. However, most existing ILI studies rely on traditional analytic methods such as ANOVA or linear mixed models, which overlook both individual differences and the inherent autocorrelation structure of time-series data. Moreover, intervention effects are often evaluated only through changes in the mean level of key variables (e.g., anxiety). This study demonstrates how dynamic structural equation modeling (DSEM) can be used to analyze ILI data and evaluate intervention effects across three dimensions—mean, autoregression, and intra-individual variability (IIV)—for two types of intervention designs: single-arm trial (SAT) and randomized controlled trial (RCT). We conducted two simulation studies to examine sample size recommendations for DSEM-based ILI studies, considering both statistical power and accuracy in parameter estimation (AIPE). In a third simulation, we compared the type I error rates of SAT and RCT designs when natural temporal changes occurred in the control group. Finally, we illustrated sample size planning using empirical data from a pre-ILI study targeting appearance anxiety reduction.

Simulation Studies 1 and 2 investigated the power and AIPE across varying sample sizes, as well as the required sample size for both SAT and RCT designs. The effect sizes of intervention effects for mean, autoregression and IIV were fixed at the medium level. Two factors regarding sample size were manipulated: number of participants (N = 30, 60, 100,150, 200, 300,400), number of time-points (T = 10, 20, 40, 60, 80, 100). The data-generating models and fitted models were identical, with analysis conducted using Mplus 8.10 and Bayesian estimation. Model performance was assessed in terms of convergence rate, power and AIPE for intervention effects, as well as bias in the standard errors of the intervention effects. Simulation Study 3 assessed the type I error rate for both designs when changes in the control group were different from zero, indicating a change (on average) due to time. Last, the empirical study conducted sample size planning based on a pre-study aimed at reducing appearance anxiety using an ILI design.

The results are as following. First, all simulation conditions achieved satisfactory convergence. Second, statistical power increased and credible interval width decreased with larger N or T. However, a minimum of 60 participants was required to achieve adequate power (i.e., ≥ 0.8). The relative bias in intervention effect was generally small. Except in the SAT design, the intervention effects on autoregression and IIV were underestimated when the number of time-points was low (T = 10 or 20). While in the RCT design, the intervention effect on mean was underestimated when sample size in both levels were small (N = 30 or 60, T = 10). Bias in the standard error was also negligible across conditions. Third, a credible interval width contours plot were useful for determining sample size under both power- and AIPE-based criteria. were useful for determining sample size under both power- and AIPE-based criteria. Fourth, when natural mean-level changes occurred between pre- and post-intervention phases, the SAT design exhibited inflated type I error rates compared to the RCT design, especially with larger samples.

In conclusion, DSEM provides a flexible framework for analyzing ILI data by simultaneously capturing intervention effects on mean, autoregression, and IIV. Researchers should choose between SAT and RCT designs

based on theoretical and practical considerations: RCTs offer stronger control for time-related confounds but require larger samples, whereas SATs are more suitable for small-sample or preliminary studies. For Monte Carlo-based sample size planning, accurate specification of true parameter values is critical; these should be derived from pre-studies, similar empirical data, or meta-analytic evidence whenever possible. When such information is unavailable, the procedures described in this study offer practical guidance.

Keywords intensive longitudinal intervention, dynamic structural equation modeling, power analysis, effect size, sample size planning

附录

2015-2025 年密集追踪干预文献检索说明

文献检索通过 X-Mol 学术平台 (https://www.x-mol.com/) 进行。X-Mol 是一个综合性学术搜索引擎, 它整合并实时更新来自包括 PubMed, Web of Science Core Collection, Scopus, Crossref, arXiv, 以及 Elsevier, Springer, Nature, Wiley, Taylor & Francis 等主要出版社期刊库在内的全球学术资源。由于密集追踪干预在文献中多定义为 “生态瞬时干预”, 因此检索聚焦于这一核心概念。采用的关键词为“ecological momentary intervention”及其标准缩写“EMI”。检索时间范围为 2015 年~2025 年。文献来源期刊信息如附表 1 所示。

附表 1　2015-2025 年密集追踪干预(生态瞬时干预)研究来源期刊情况

期刊名称	文献篇数	是否 SSCI	是否 SCIE	是否 ESCI	JCR 分区	中科院分区
Internet Interventions (IF 3.6)	3		√		Q1	2 区
PSYCHIATRY RESEARCH (IF 3.9)	1		√		Q1	3 区
JAMA Psychiatry (IF 22.5)	1		√		Q1	1 区
Journal of Anxiety Disorders (IF 4.8)	1	√			Q1	2 区
JMIR Mental Health (IF 4.8)	3		√		Q1	2 区
Journal of Medical Internet Research (IF 5.8)	3		√		Q1	2 区
Journal of Consulting and Clinical Psychology (IF 4.5)	1	√			Q1	1 区
Schizophrenia Bulletin (IF 5.3)	1		√		Q1	1 区
BMC Public Health (IF 3.5)	1		√		Q1	2 区
Applied Psychology: Health and Well-Being (IF 3.8)	3	√			Q1	2 区
JMIR Formative Research (IF 2.0)	1		√		Q3	4 区
Collabra: Psychology (IF 3.1)	1	√			Q1	3 区
Informatics for Health & Social Care (IF 2.5)	1		√		Q2	4 区
Journal of Counseling Psychology (IF 3.8)	1	√			Q1	1 区
European Psychiatry (IF 7.2)	1		√		Q1	2 区
Trials (IF 2.0)	3		√		Q3	4 区
Cyberpsychology, Behavior, and Social Networking (IF 4.2)	1	√			Q1	3 区
JMIR mHealth and uHealth (IF 5.4)	1		√		Q1	2 区
Depression and Anxiety (IF 4.7)	1		√		Q1	2 区
Psychological Medicine (IF 5.9)	1		√		Q1	2 区
Schizophrenia Research (IF 3.6)	1		√		Q1	2 区
JMIR Research Protocols	1			√	Q3	4 区
Psychol Methods (IF 7.8)	1	√			Q1	1 区
Affective Science (IF 2.6)	1			√	Q2	4 区
四川师范大学硕士学位论文集	1					
华中师范大学硕士学位论文集	1					

附表 2　单臂设计下非平衡设计的收敛率

N	*T*					
	20	40	80	120	160	200
30	49%	100%	100%	100%	100%	100%
60	100%	100%	100%	100%	99%	99%
100	100%	100%	99%	99%	99%	99%
150	100%	100%	100%	99%	99%	99%
200	100%	100%	100%	100%	99%	100%
300	100%	100%	100%	100%	100%	100%
400	99%	99%	100%	99%	100%	100%

附表 3 单臂设计非平衡设计情况下检验力、相对偏差、可信区间宽度结果

N	T	power			rbias			width		
		γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}
30	20	**0.106**	**0.176**	**0.102**	0.070	−0.074	**0.119**	0.559	0.437	0.569
	40	**0.164**	**0.270**	**0.150**	0.057	−0.011	0.012	0.424	0.305	0.422
	80	**0.332**	**0.498**	**0.262**	−0.020	−0.014	0.009	0.278	0.216	0.300
	120	**0.464**	**0.680**	**0.360**	0.025	0.021	−0.063	0.239	0.184	0.244
	160	**0.531**	**0.768**	**0.515**	−0.062	−0.037	0.027	0.208	0.160	0.216
	200	**0.640**	0.882	**0.632**	−0.014	0.025	−0.003	0.187	0.145	0.194
60	20	**0.192**	**0.300**	**0.162**	−0.034	−0.082	**−0.114**	0.374	0.301	0.395
	40	**0.436**	**0.538**	**0.258**	**0.104**	−0.060	−0.076	0.259	0.210	0.295
	80	**0.614**	0.850	**0.598**	0.026	−0.002	0.027	0.193	0.147	0.200
	120	**0.798**	0.944	**0.716**	0.033	−0.021	−0.060	0.163	0.119	0.165
	160	0.877	0.982	0.859	−0.025	−0.005	0.016	0.140	0.112	0.144
	200	0.933	0.996	0.899	−0.025	0.004	−0.022	0.125	0.100	0.135
100	20	**0.316**	**0.412**	**0.196**	−0.079	−0.100	**−0.197**	0.275	0.234	0.317
	40	**0.562**	**0.754**	**0.430**	−0.034	−0.067	−0.062	0.200	0.161	0.225
	80	0.885	0.978	**0.780**	0.022	0.001	0.008	0.152	0.111	0.152
	120	0.970	0.994	0.889	0.007	−0.014	−0.025	0.122	0.092	0.133
	160	0.978	1.000	0.968	−0.040	0.015	−0.005	0.107	0.082	0.112
	200	0.990	1.000	0.982	−0.019	0.004	−0.023	0.099	0.075	0.103
150	20	**0.568**	**0.740**	**0.444**	0.008	−0.086	**−0.146**	0.211	0.158	0.210
	40	**0.780**	0.942	**0.662**	−0.002	0.021	−0.060	0.157	0.121	0.175
	80	0.986	0.998	0.954	−0.017	0.003	0.004	0.110	0.084	0.114
	120	1.000	1.000	0.992	0.000	−0.002	−0.013	0.092	0.069	0.096
	160	1.000	1.000	0.996	−0.012	0.000	0.000	0.083	0.064	0.088
	200	1.000	1.000	1.000	0.001	0.007	−0.002	0.076	0.060	0.076
200	20	**0.706**	**0.734**	**0.454**	0.041	−0.049	**−0.101**	0.196	0.165	0.217
	40	0.884	0.970	**0.776**	0.018	−0.020	−0.024	0.139	0.103	0.163
	80	0.996	1.000	0.984	−0.011	−0.001	0.030	0.098	0.082	0.106
	120	1.000	1.000	1.000	−0.001	−0.026	−0.001	0.088	0.065	0.089
	160	1.000	1.000	1.000	−0.018	0.004	0.007	0.068	0.053	0.069
	200	1.000	1.000	1.000	−0.010	0.000	−0.025	0.063	0.049	0.066
300	20	0.808	0.880	**0.702**	0.024	−0.056	−0.046	0.169	0.134	0.166
	40	0.954	0.998	0.930	−0.022	0.004	−0.001	0.114	0.092	0.127
	80	1.000	1.000	1.000	−0.009	0.002	0.063	0.080	0.063	0.084
	120	1.000	1.000	1.000	0.017	−0.009	0.000	0.066	0.053	0.072
	160	1.000	1.000	1.000	−0.012	0.004	−0.003	0.061	0.046	0.064
	200	1.000	1.000	1.000	−0.001	0.005	−0.002	0.055	0.043	0.059
400	20	0.901	0.978	0.873	−0.017	−0.063	−0.086	0.123	0.092	0.126
	40	0.996	1.000	0.992	−0.012	−0.003	−0.005	0.086	0.066	0.094
	80	1.000	1.000	1.000	−0.013	0.004	−0.012	0.065	0.049	0.067
	120	1.000	1.000	1.000	0.007	−0.003	−0.008	0.054	0.042	0.057
	160	1.000	1.000	1.000	−0.012	0.002	0.003	0.047	0.039	0.051
	200	1.000	1.000	1.000	−0.009	0.000	−0.007	0.044	0.036	0.048

注：N 表示被试量, T 表示测量时间点总数量, power 表示检验力, 其中加粗的表示小于 0.8 的结果, rbias 表示相对偏差, 其中加粗的表示在[−0.1,0.1]范围外的结果, width 表示可信区间宽度。

附表 4　单臂设计平衡设计情况下误差均方根、覆盖率、标准误偏差结果

N	T	RMSE			Coverage			SE−SD		
		γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}
30	20	0.121	0.094	0.135	**0.976**	0.960	0.960	0.015	0.010	0.007
	40	0.086	0.063	0.089	**0.976**	0.962	0.972	0.012	0.007	0.008
	80	0.064	0.044	0.061	0.968	0.970	0.962	0.005	0.004	0.008
	120	0.052	0.035	0.051	0.960	0.964	0.970	0.005	0.005	0.005
	160	0.044	0.035	0.047	**0.978**	0.952	0.958	0.006	0.001	0.003
	200	0.040	0.030	0.042	0.966	0.958	0.970	0.006	0.003	0.003
60	20	0.086	0.067	0.094	0.962	0.958	0.956	0.005	0.004	0.004
	40	0.059	0.045	0.058	**0.978**	0.950	**0.976**	0.006	0.003	0.008
	80	0.044	0.032	0.045	0.962	0.974	0.962	0.002	0.001	0.001
	120	0.037	0.025	0.035	0.948	0.966	0.972	0.001	0.002	0.003
	160	0.032	0.025	0.031	0.960	0.944	0.966	0.001	0.000	0.003
	200	0.028	0.021	0.030	0.962	0.968	0.952	0.002	0.001	0.000
100	20	0.064	0.053	0.070	0.974	0.946	0.960	0.005	0.001	0.004
	40	0.048	0.035	0.049	0.956	0.954	0.964	0.001	0.002	0.002
	80	0.035	0.025	0.035	0.948	0.952	0.958	0.000	0.000	0.000
	120	0.029	0.020	0.029	0.958	0.964	0.956	0.000	0.001	0.000
	160	0.024	0.019	0.023	0.960	0.940	0.968	0.001	−0.001	0.002
	200	0.022	0.017	0.023	0.958	0.950	0.940	0.002	0.000	0.000
150	20	0.053	0.043	0.059	0.950	0.952	0.954	0.003	0.001	0.001
	40	0.039	0.030	0.040	0.944	0.956	0.956	0.000	−0.001	0.001
	80	0.028	0.022	0.029	0.946	0.946	0.942	−0.001	−0.001	0.000
	120	0.023	0.018	0.024	0.941	0.927	0.957	0.000	−0.001	−0.001
	160	0.021	0.016	0.021	0.939	0.937	0.957	−0.001	−0.001	0.000
	200	0.018	0.014	0.018	0.952	0.939	0.962	0.000	0.000	0.001
200	20	0.044	0.036	0.049	0.966	0.950	0.958	0.004	0.002	0.003
	40	0.034	0.025	0.035	0.948	0.956	0.962	−0.001	0.001	0.000
	80	0.024	0.018	0.024	0.944	0.952	0.966	0.000	0.000	0.000
	120	0.020	0.015	0.021	0.948	0.936	0.940	0.000	0.000	−0.001
	160	0.017	0.014	0.017	**0.978**	0.934	0.948	0.001	−0.001	0.001
	200	0.015	0.012	0.016	0.974	0.956	0.948	0.001	0.000	0.000
300	20	0.038	0.030	0.043	0.956	0.954	0.962	0.001	0.000	0.000
	40	0.028	0.021	0.029	0.930	0.936	0.954	−0.001	−0.001	−0.001
	80	0.021	0.014	0.021	0.942	0.954	0.948	−0.001	0.000	−0.001
	120	0.016	0.013	0.017	0.962	0.938	0.952	0.000	−0.001	−0.001
	160	0.014	0.011	0.014	0.950	0.938	0.962	0.000	0.000	0.000
	200	0.012	0.010	0.013	0.974	0.960	0.944	0.001	0.000	0.000
400	20	0.031	0.028	0.036	0.972	**0.910**	0.962	0.003	−0.002	0.000
	40	0.024	0.018	0.026	0.944	0.934	0.944	−0.001	−0.001	−0.001
	80	0.017	0.013	0.019	0.948	0.942	0.950	0.000	0.000	−0.001
	120	0.014	0.011	0.015	0.944	**0.924**	0.946	0.000	−0.001	0.000
	160	0.012	0.010	0.012	0.950	0.932	0.954	0.000	−0.001	0.000
	200	0.010	0.008	0.011	0.964	0.950	0.960	0.001	0.000	0.000

注：N表示被试量, T表示测量时间点总数量, RMSE 表示误差均方根, Coverage 表示可信区间覆盖率, 其中加粗的表示在[0.925, 0.975]范围外的结果, SE-SD 表示估计标准误相对于其估计值标准差的偏差。

附表 5 单臂设计非平衡设计情况下误差均方根、覆盖率、标准误偏差结果

N	T	RMSE			Coverage			SE−SD		
		γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}	γ_{30}	γ_{40}	ω_{50}
30	20	0.126	0.110	0.143	0.967	0.955	0.935	0.015	0.001	0.004
	40	0.093	0.076	0.096	**0.978**	0.954	0.974	0.015	0.001	0.011
	80	0.064	0.050	0.070	0.966	0.962	**0.978**	0.006	0.005	0.007
	120	0.054	0.044	0.060	0.972	0.968	0.958	0.006	0.003	0.003
	160	0.050	0.037	0.053	0.964	0.960	0.956	0.003	0.004	0.003
	200	0.042	0.034	0.046	**0.976**	0.960	0.960	0.006	0.003	0.004
60	20	0.091	0.077	0.100	0.954	0.948	0.944	0.004	0.000	0.003
	40	0.064	0.050	0.069	0.942	0.958	0.970	0.003	0.004	0.007
	80	0.047	0.036	0.047	0.960	0.952	0.964	0.002	0.002	0.004
	120	0.039	0.031	0.041	0.966	0.966	0.962	0.002	0.000	0.001
	160	0.034	0.026	0.038	0.958	0.960	0.933	0.001	0.003	−0.001
	200	0.031	0.024	0.032	0.939	0.970	0.958	0.001	0.002	0.003
100	20	0.072	0.061	0.083	0.944	0.950	0.950	0.000	−0.001	0.001
	40	0.047	0.039	0.054	0.960	0.956	0.960	0.004	0.003	0.004
	80	0.034	0.028	0.039	**0.976**	0.947	0.939	0.004	0.000	−0.001
	120	0.029	0.024	0.035	0.964	0.954	0.945	0.002	0.000	−0.001
	160	0.026	0.021	0.029	0.964	0.945	0.941	0.002	0.000	−0.001
	200	0.025	0.019	0.026	0.962	0.952	0.941	0.000	0.000	0.001
150	20	0.051	0.042	0.059	**0.976**	0.938	0.930	0.003	0.000	−0.003
	40	0.042	0.032	0.043	0.938	0.944	0.952	−0.003	−0.001	0.002
	80	0.026	0.021	0.031	0.970	0.966	0.936	0.002	0.000	−0.001
	120	0.022	0.019	0.024	0.964	0.942	0.956	0.001	−0.001	0.000
	160	0.020	0.017	0.022	0.952	0.938	0.956	0.001	0.000	0.000
	200	0.019	0.015	0.021	0.960	0.949	0.925	0.000	0.000	−0.002
200	20	0.049	0.043	0.059	0.940	0.946	0.942	0.002	0.000	−0.002
	40	0.036	0.027	0.038	0.956	0.936	0.966	−0.001	−0.001	0.003
	80	0.025	0.020	0.029	0.958	0.960	0.938	0.001	0.000	−0.002
	120	0.020	0.017	0.024	0.964	0.934	0.942	0.002	0.000	−0.001
	160	0.018	0.014	0.019	0.941	0.954	0.929	0.000	0.000	−0.001
	200	0.016	0.012	0.018	0.954	0.948	0.940	0.000	0.000	−0.001
300	20	0.040	0.036	0.048	0.960	0.932	**0.902**	0.003	−0.001	−0.005
	40	0.030	0.024	0.032	0.938	0.948	0.950	−0.001	0.000	0.001
	80	0.021	0.017	0.024	0.958	0.936	0.930	0.000	−0.001	−0.001
	120	0.018	0.014	0.020	0.948	0.946	0.932	0.000	0.000	−0.002
	160	0.015	0.013	0.016	0.972	**0.924**	0.954	0.001	−0.001	0.001
	200	0.014	0.011	0.016	0.956	0.960	0.932	0.000	0.000	−0.001
400	20	0.031	0.027	0.034	0.956	**0.923**	0.942	0.000	−0.002	−0.001
	40	0.022	0.017	0.023	0.948	0.946	0.950	0.000	−0.001	0.001
	80	0.016	0.013	0.018	0.942	0.938	0.938	0.000	−0.001	−0.001
	120	0.014	0.011	0.016	0.950	0.958	0.946	0.000	0.000	−0.001
	160	0.012	0.010	0.014	0.952	0.952	0.944	0.000	−0.001	−0.001
	200	0.012	0.009	0.012	0.952	0.974	0.960	0.000	0.000	0.000

注：N表示被试量, T表示测量时间点总数量, RMSE 表示误差均方根, Coverage 表示可信区间覆盖率, 其中加粗的表示在[0.925, 0.975]范围外的结果, SE-SD 表示估计标准误相对于其估计值标准差的偏差。

附表 6　随机对照设计非平衡设计下检验力、相对偏差、可信区间宽度结果

N	T	power			rbias			width		
		γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}
30	20	**0.062**	**0.086**	**0.108**	**−0.114**	−0.073	0.085	0.824	0.581	0.773
	40	**0.124**	**0.186**	**0.134**	0.027	0.001	−0.006	0.598	0.455	0.578
	80	**0.180**	**0.278**	**0.172**	0.015	−0.072	−0.039	0.461	0.353	0.461
	120	**0.228**	**0.334**	**0.278**	−0.012	−0.048	0.030	0.409	0.320	0.398
	160	**0.274**	**0.456**	**0.276**	−0.052	0.013	−0.021	0.367	0.291	0.363
	200	**0.336**	**0.528**	**0.342**	−0.013	0.008	0.004	0.345	0.268	0.348
60	20	**0.158**	**0.282**	**0.156**	−0.043	−0.030	−0.068	0.506	0.376	0.497
	40	**0.290**	**0.454**	**0.270**	0.031	0.018	−0.016	0.389	0.298	0.377
	80	**0.442**	**0.606**	**0.428**	−0.029	−0.052	−0.033	0.307	0.237	0.297
	120	**0.548**	**0.742**	**0.538**	−0.005	0.005	−0.007	0.266	0.210	0.263
	160	**0.578**	0.806	**0.608**	−0.033	−0.004	−0.019	0.245	0.194	0.242
	200	**0.712**	0.862	**0.696**	−0.009	0.004	0.016	0.223	0.181	0.229
100	20	**0.258**	**0.464**	**0.288**	−0.077	−0.032	−0.002	0.377	0.286	0.374
	40	**0.442**	**0.720**	**0.480**	−0.009	0.020	−0.039	0.295	0.223	0.279
	80	**0.676**	0.858	**0.726**	−0.001	−0.027	0.021	0.228	0.180	0.225
	120	0.802	0.936	0.818	0.034	−0.009	0.004	0.199	0.159	0.196
	160	0.866	0.958	0.850	−0.013	−0.017	−0.015	0.181	0.147	0.182
	200	0.902	0.974	0.904	0.001	0.003	0.005	0.169	0.139	0.170
150	20	**0.404**	**0.698**	**0.488**	−0.048	0.013	0.040	0.303	0.227	0.297
	40	**0.648**	0.876	**0.696**	−0.013	0.019	0.020	0.234	0.178	0.231
	80	0.874	0.960	0.890	0.009	−0.029	0.034	0.183	0.146	0.181
	120	0.934	0.994	0.954	0.026	−0.006	−0.008	0.160	0.127	0.156
	160	0.964	0.998	0.950	−0.001	0.000	−0.006	0.144	0.120	0.147
	200	0.980	0.996	0.986	−0.004	0.003	0.013	0.136	0.111	0.136
200	20	**0.566**	0.800	**0.570**	−0.008	−0.007	0.015	0.258	0.196	0.255
	40	**0.796**	0.942	0.810	0.020	0.015	0.005	0.200	0.154	0.195
	80	0.930	0.994	0.954	−0.011	−0.028	0.019	0.155	0.126	0.153
	120	0.976	0.998	0.978	0.018	−0.019	0.011	0.138	0.111	0.135
	160	0.992	1.000	0.998	−0.012	0.000	−0.004	0.126	0.101	0.126
	200	0.998	1.000	0.992	−0.003	0.001	0.001	0.117	0.096	0.118
300	20	**0.740**	0.946	**0.782**	−0.025	0.007	0.012	0.212	0.156	0.206
	40	0.934	0.994	0.938	−0.003	−0.002	−0.002	0.156	0.124	0.157
	80	0.986	1.000	0.996	−0.010	−0.018	0.026	0.124	0.100	0.126
	120	0.998	1.000	0.994	0.013	−0.018	0.008	0.110	0.090	0.109
	160	1.000	1.000	1.000	−0.011	−0.001	0.007	0.101	0.083	0.103
	200	1.000	1.000	1.000	−0.006	−0.004	0.012	0.095	0.078	0.095
400	20	0.866	0.992	0.866	−0.002	0.001	0.008	0.176	0.133	0.177
	40	0.982	1.000	0.984	0.023	0.008	0.035	0.138	0.106	0.137
	80	1.000	1.000	0.996	0.005	−0.009	0.011	0.110	0.086	0.109
	120	1.000	1.000	1.000	0.009	−0.014	0.002	0.096	0.077	0.095
	160	1.000	1.000	1.000	−0.015	−0.001	0.007	0.086	0.072	0.088
	200	1.000	1.000	1.000	0.003	0.001	−0.001	0.082	0.067	0.082

注：N 表示被试量(N/2 表示每组被试量), T 表示测量时间点总数量, power 表示检验力, 其中加粗的表示小于 0.8 的结果, rbias 表示相对偏差, 其中加粗的表示在[−0.1,0.1]范围外的结果, width 表示可信区间宽度。

附表 7 随机对照设计平衡设计下误差均方根、覆盖率、标准误偏差结果

N	T	RMSE			Coverage			SE−SD		
		γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}
30	20	0.201	0.137	0.191	**0.994**	**0.990**	**0.992**	0.036	0.035	0.045
	40	0.140	0.106	0.148	**0.984**	**0.980**	**0.980**	0.032	0.016	0.020
	80	0.111	0.084	0.109	**0.976**	0.962	**0.978**	0.015	0.008	0.016
	120	0.096	0.070	0.096	0.974	0.962	0.972	0.011	0.009	0.011
	160	0.083	0.066	0.086	0.974	0.972	0.962	0.012	0.005	0.009
	200	0.073	0.062	0.080	**0.976**	0.960	0.964	0.014	0.005	0.007
60	20	0.134	0.098	0.135	**0.976**	0.974	**0.976**	0.018	0.015	0.017
	40	0.096	0.078	0.104	**0.980**	0.966	0.962	0.014	0.003	0.005
	80	0.073	0.056	0.079	**0.982**	0.956	0.958	0.009	0.005	0.003
	120	0.067	0.050	0.067	0.950	0.972	0.960	0.002	0.003	0.003
	160	0.059	0.048	0.062	0.963	0.949	0.943	0.004	0.000	0.001
	200	0.051	0.043	0.055	0.974	0.949	0.966	0.006	0.001	0.003
100	20	0.104	0.080	0.108	0.964	0.956	0.966	0.009	0.005	0.005
	40	0.075	0.060	0.081	0.970	0.966	0.956	0.008	0.001	0.001
	80	0.055	0.044	0.058	**0.978**	0.962	0.968	0.007	0.002	0.004
	120	0.050	0.040	0.051	0.948	0.946	0.946	0.002	0.000	0.002
	160	0.046	0.036	0.046	0.955	0.945	0.959	0.001	0.000	0.001
	200	0.041	0.032	0.042	0.953	0.972	0.961	0.003	0.001	0.001
150	20	0.084	0.064	0.087	0.970	0.956	0.970	0.007	0.003	0.003
	40	0.057	0.049	0.064	**0.984**	0.954	0.964	0.009	0.000	0.002
	80	0.046	0.036	0.048	0.964	0.946	0.966	0.004	0.001	0.001
	120	0.042	0.031	0.042	0.946	0.974	0.964	0.000	0.001	0.001
	160	0.036	0.029	0.038	0.956	0.956	0.948	0.001	0.001	0.000
	200	0.034	0.026	0.035	0.950	0.958	0.948	0.000	0.001	0.001
200	20	0.072	0.056	0.075	0.968	0.942	0.962	0.006	0.001	0.003
	40	0.050	0.042	0.057	**0.980**	0.950	0.956	0.007	0.000	0.000
	80	0.039	0.031	0.040	0.966	0.948	0.958	0.003	0.001	0.002
	120	0.035	0.027	0.037	0.956	0.958	0.948	0.001	0.000	0.000
	160	0.032	0.025	0.031	0.948	0.948	0.964	0.001	0.000	0.001
	200	0.029	0.023	0.029	0.958	0.950	0.948	0.000	0.001	0.001
300	20	0.060	0.048	0.062	0.960	0.936	0.966	0.003	−0.002	0.000
	40	0.044	0.031	0.043	0.964	0.968	0.974	0.002	0.003	0.002
	80	0.034	0.026	0.034	0.953	0.947	0.957	0.000	−0.001	0.000
	120	0.029	0.021	0.031	0.942	0.964	0.938	0.001	0.002	−0.001
	160	0.026	0.020	0.027	0.950	0.956	0.946	0.000	0.000	−0.001
	200	0.024	0.019	0.024	0.957	0.966	0.962	0.000	0.000	0.001
400	20	0.053	0.041	0.054	0.958	0.944	0.946	0.001	−0.001	−0.001
	40	0.037	0.028	0.041	0.964	0.958	0.954	0.003	0.002	−0.001
	80	0.029	0.023	0.029	0.960	0.940	0.952	0.001	0.000	0.000
	120	0.025	0.019	0.026	0.956	0.950	0.954	0.000	0.000	−0.001
	160	0.023	0.018	0.023	0.946	0.956	0.950	0.000	0.000	0.000
	200	0.021	0.017	0.020	0.930	0.958	0.960	0.000	−0.001	0.001

注：N表示被试量, T表示测量时间点总数量, RMSE 表示误差均方根, Coverage 表示可信区间覆盖率, 其中加粗的表示在[0.925, 0.975]范围外的结果, SE-SD 表示估计标准误相对于其估计值标准差的偏差。

附表 8　随机对照设计非平衡设计下误差均方根、覆盖率、标准误偏差结果

N	T	RMSE			Coverage			SE-SD		
		γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}
30	20	0.177	0.122	0.180	0.972	**0.978**	0.970	0.032	0.026	0.016
	40	0.136	0.100	0.122	0.968	**0.976**	**0.978**	0.017	0.015	0.024
	80	0.104	0.078	0.102	0.966	**0.978**	0.972	0.013	0.012	0.015
	120	0.092	0.074	0.094	0.962	0.974	0.962	0.012	0.007	0.007
	160	0.083	0.066	0.080	0.970	0.960	0.966	0.010	0.008	0.012
	200	0.078	0.064	0.077	0.968	0.964	0.970	0.010	0.004	0.011
60	20	0.121	0.085	0.119	0.952	0.972	0.952	0.009	0.011	0.007
	40	0.094	0.072	0.087	0.968	0.960	**0.976**	0.005	0.004	0.009
	80	0.073	0.055	0.073	0.954	0.960	0.946	0.005	0.005	0.003
	120	0.064	0.052	0.062	0.958	0.958	0.956	0.003	0.001	0.005
	160	0.058	0.048	0.058	0.966	0.968	0.956	0.004	0.002	0.003
	200	0.054	0.045	0.054	0.962	0.948	0.972	0.003	0.001	0.004
100	20	0.089	0.063	0.089	0.958	**0.978**	0.954	0.007	0.010	0.006
	40	0.072	0.055	0.070	0.956	0.960	0.948	0.002	0.002	0.002
	80	0.054	0.044	0.054	0.974	0.950	0.962	0.004	0.002	0.003
	120	0.050	0.039	0.049	0.958	0.950	0.940	0.001	0.001	0.000
	160	0.043	0.037	0.046	0.962	0.958	0.952	0.004	0.000	0.001
	200	0.042	0.035	0.041	0.960	0.944	0.968	0.001	0.000	0.002
150	20	0.074	0.052	0.072	0.954	0.974	0.962	0.003	0.006	0.003
	40	0.059	0.045	0.055	0.954	0.956	0.958	0.001	0.001	0.004
	80	0.045	0.036	0.045	0.962	0.952	0.956	0.002	0.001	0.002
	120	0.041	0.031	0.038	0.948	0.956	0.948	0.000	0.001	0.002
	160	0.038	0.030	0.036	0.948	0.964	0.952	–0.001	0.001	0.001
	200	0.033	0.029	0.034	0.942	0.940	0.960	0.001	0.000	0.000
200	20	0.062	0.046	0.063	0.960	0.974	0.968	0.003	0.004	0.002
	40	0.049	0.040	0.049	0.954	0.948	0.954	0.002	–0.001	0.001
	80	0.038	0.030	0.038	0.956	0.954	0.956	0.001	0.002	0.001
	120	0.037	0.027	0.034	0.932	0.958	0.948	–0.002	0.001	0.000
	160	0.032	0.026	0.030	0.952	0.964	0.964	0.000	0.000	0.002
	200	0.029	0.024	0.030	0.958	0.968	0.952	0.000	0.001	0.000
300	20	0.052	0.036	0.051	0.948	0.970	0.948	0.003	0.004	0.001
	40	0.042	0.030	0.040	0.944	0.960	0.948	–0.002	0.001	0.000
	80	0.031	0.024	0.032	0.954	0.956	0.948	0.001	0.002	0.000
	120	0.030	0.022	0.028	0.938	0.966	0.940	–0.002	0.001	0.000
	160	0.025	0.021	0.025	0.944	0.968	0.966	0.000	0.000	0.001
	200	0.023	0.020	0.024	0.962	0.952	0.952	0.001	0.000	0.000
400	20	0.045	0.033	0.044	0.956	0.962	0.958	0.000	0.001	0.001
	40	0.035	0.028	0.035	0.942	0.940	0.958	0.000	–0.001	0.000
	80	0.027	0.022	0.028	0.964	0.948	0.956	0.001	0.000	0.000
	120	0.025	0.020	0.024	0.936	0.950	0.946	–0.001	0.000	0.000
	160	0.022	0.019	0.022	0.942	0.944	0.958	0.000	0.000	0.000
	200	0.021	0.017	0.020	0.962	0.960	0.954	0.000	0.000	0.001

注：N表示被试量, T表示测量时间点总数量, RMSE表示误差均方根, Coverage表示可信区间覆盖率, 其中加粗的表示在[0.925, 0.975]范围外的结果, SE-SD 表示估计标准误相对于其估计值标准差的偏差。

附表 9 基于预实验结果的随机对照设计下检验力、相对偏差、可信区间宽度结果

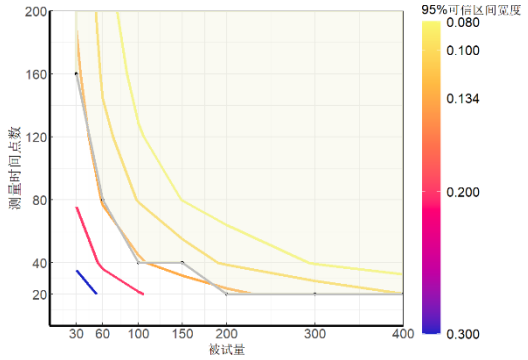
N	T	power			rbias			width		
		γ_{31}	γ_{41}	ω_{11}	γ_{31}	γ_{41}	ω_{11}	γ_{31}	γ_{41}	ω_{11}
200	40	**0.690**	**0.060**	**0.712**	0.001	−0.057	0.000	0.245	0.199	0.528
	80	**0.728**	**0.078**	**0.786**	−0.006	**−0.151**	0.013	0.235	0.161	0.490
	120	**0.752**	**0.116**	0.826	0.024	0.025	0.013	0.231	0.147	0.477
	160	**0.724**	**0.094**	0.810	−0.002	−0.035	0.009	0.229	0.139	0.473
	200	**0.766**	**0.098**	0.840	0.019	0.013	0.005	0.228	0.134	0.467
250	40	**0.758**	**0.072**	0.802	0.004	**−0.126**	0.007	0.217	0.177	0.473
	80	0.814	**0.082**	0.872	−0.003	**−0.107**	0.013	0.209	0.143	0.438
	120	0.848	**0.116**	0.878	0.022	−0.007	0.007	0.206	0.131	0.427
	160	0.836	**0.090**	0.900	0.000	−0.015	0.016	0.204	0.124	0.422
	200	0.864	**0.124**	0.902	0.026	0.000	−0.002	0.203	0.120	0.417
300	40	0.850	**0.082**	0.886	−0.003	**−0.133**	0.004	0.199	0.162	0.432
	80	0.894	**0.108**	0.920	−0.003	−0.029	0.008	0.191	0.131	0.400
	120	0.928	**0.144**	0.920	0.028	0.006	0.004	0.188	0.119	0.389
	160	0.908	**0.110**	0.932	−0.002	−0.054	0.004	0.186	0.113	0.385
	200	0.912	**0.140**	0.950	0.019	−0.003	−0.003	0.185	0.109	0.381
400	40	0.930	**0.072**	0.944	0.003	**−0.126**	0.000	0.172	0.140	0.375
	80	0.962	**0.118**	0.976	0.006	−0.014	0.004	0.165	0.113	0.346
	120	0.970	**0.158**	0.982	0.021	0.005	0.010	0.162	0.103	0.338
	160	0.970	**0.134**	0.982	0.007	−0.079	0.000	0.160	0.098	0.334
	200	0.984	**0.178**	0.980	0.028	0.035	−0.007	0.160	0.094	0.330

注：N 表示被试量, T 表示测量时间点总数量, power 表示检验力, 其中加粗的表示小于 0.8 的结果, rbias 表示相对偏差, 其中加粗的表示在[−0.1,0.1]范围外的结果, width 表示可信区间宽度。

附表 10 基于预实验结果的随机对照设计下误差均方根、覆盖率、标准误偏差结果

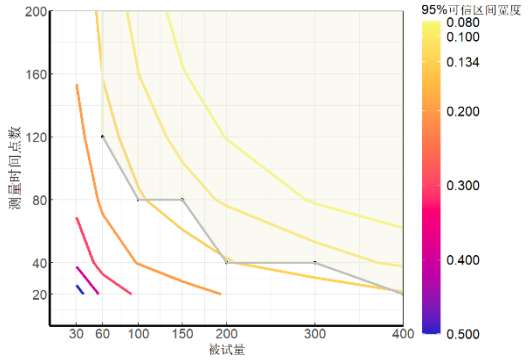
N	T	RMSE			Coverage			SE−SD		
		γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}	γ_{03}	γ_{41}	ω_{05}
200	40	0.066	0.050	0.132	0.944	0.954	0.958	−0.003	0.000	0.000
	80	0.059	0.041	0.124	0.948	0.948	0.942	0.000	−0.001	−0.001
	120	0.057	0.040	0.118	0.960	0.930	0.966	0.002	−0.003	0.001
	160	0.059	0.034	0.122	0.946	0.964	0.952	−0.001	0.002	−0.004
	200	0.055	0.035	0.111	0.956	0.944	0.962	0.003	−0.001	0.006
250	40	0.060	0.046	0.117	0.932	0.950	0.956	−0.004	−0.001	0.001
	80	0.055	0.037	0.107	0.930	0.954	0.968	−0.002	0.000	0.003
	120	0.050	0.034	0.107	0.956	0.948	0.956	0.002	−0.001	−0.001
	160	0.052	0.029	0.105	0.954	0.970	0.954	0.000	0.002	0.001
	200	0.049	0.031	0.100	0.962	0.946	0.976	0.002	0.000	0.004
300	40	0.051	0.041	0.105	0.946	0.958	0.960	−0.001	0.000	0.002
	80	0.048	0.033	0.096	0.952	0.952	0.952	0.001	0.000	0.003
	120	0.046	0.032	0.099	0.958	0.950	0.952	0.002	−0.002	−0.002
	160	0.046	0.027	0.097	0.946	0.962	0.950	0.001	0.001	0.000
	200	0.046	0.028	0.091	0.964	0.954	0.962	0.002	0.000	0.004
400	40	0.044	0.034	0.087	0.956	0.964	0.970	0.000	0.001	0.006
	80	0.044	0.029	0.083	0.952	0.948	0.960	−0.002	0.000	0.003
	120	0.041	0.027	0.086	0.948	0.940	0.956	0.001	−0.001	−0.002
	160	0.041	0.024	0.083	0.954	0.962	0.956	0.000	0.001	0.000
	200	0.039	0.024	0.085	0.964	0.956	0.944	0.002	0.000	−0.003

注：N 表示被试量, T 表示测量时间点数量, RMSE 表示误差均方根, Coverage 表示可信区间覆盖率, 其中加粗的表示在[0.925, 0.975]范围外的结果, SE-SD 表示估计标准误相对于其估计值标准差的偏差。



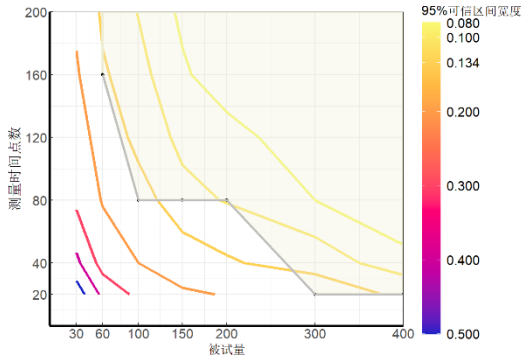
附图 1　单臂设计平衡设计下γ_{40}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。

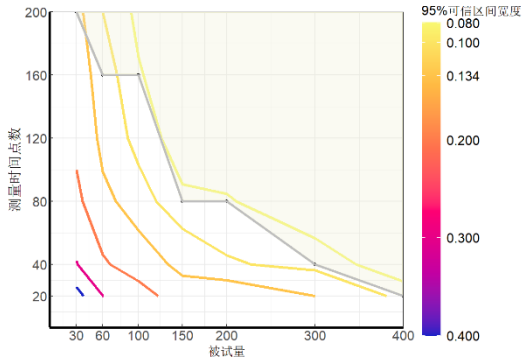


附图 2　单臂设计平衡设计下ω_{50}的可信区间宽度等高线图

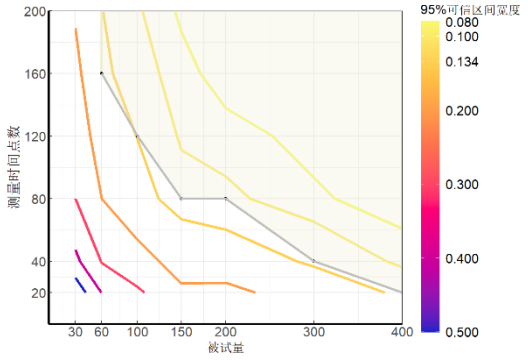
注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。



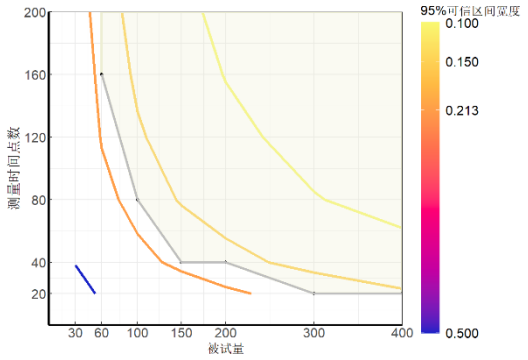
附图 3　单臂设计非平衡设计下γ_{30}的可信区间宽度等高线图



附图 4　单臂设计非平衡设计下γ_{40}的可信区间宽度等高线图

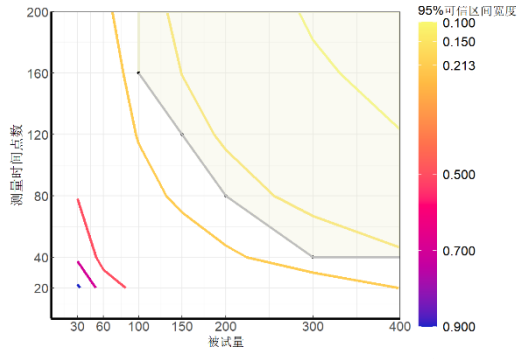


附图 5　单臂设计非平衡设计下ω_{50}的可信区间宽度等高线图



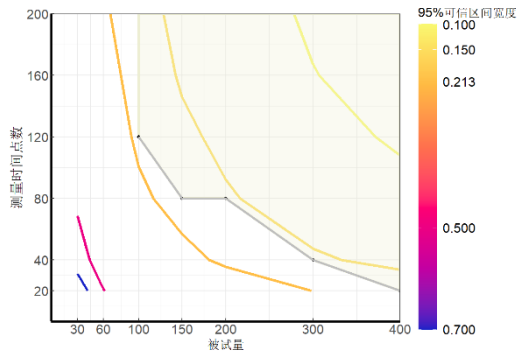
附图 6　随机对照设计平衡设计下γ_{41}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。



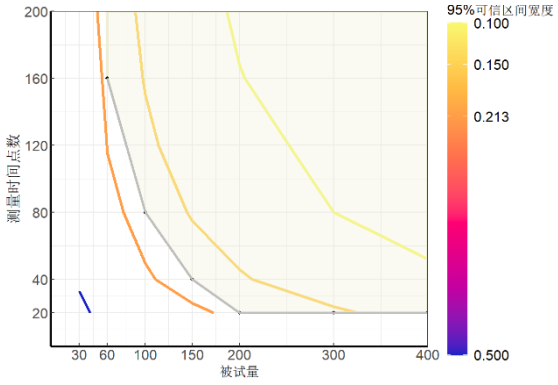
附图 7　随机对照设计平衡设计下ω_{05}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。



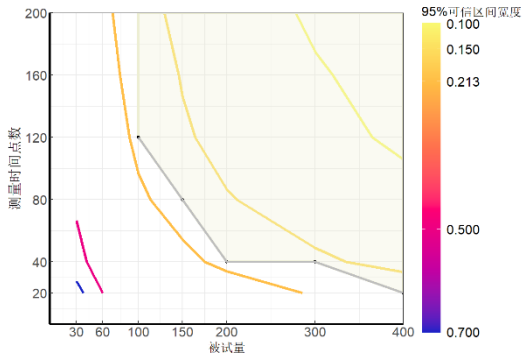
附图 8　随机对照设计非平衡设计下γ_{03}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。



附图 9　随机对照设计非平衡设计下γ_{41}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。



附图 10　随机对照设计非平衡设计下ω_{05}的可信区间宽度等高线图

注：图中阴影区域表示符合检验力大于等于 0.8 标准的条件。

社交媒体外貌焦虑生态瞬时干预的预实验流程及结果

预实验流程如下。首先，招募被试，完成资格审查问卷和前测问卷的填写。资格审查问卷主要用于筛查不符合实验要求的被试，包括外貌特质焦虑水平过高或过低的被试，近期内在手机上没有使用过社交媒体的被试等。前测问卷主要包括一些个体水平变量，例如自我同情水平等。符合预实验资格审查要求并认真完成[3]前测量表填写的共 237 人。这些被试被平均分为两组(干预组、控制组)，所有被试完成前后测各 20 天，每天 1 次的密集追踪测量，通过自我报告的方式收集每天的社交媒体外貌焦虑水平(1-100 点计分)。在后测 20 天中，干预组每天需完成自我同情相关材料[4]的阅读，控制组需完成中性材料的阅读。实验完成后，去掉中途流失的被试，最终保留 200 名被试，其中干预组 105 人，控制组 95 人。

以社交媒体外貌焦虑作为结果变量，采用模型 3 分析数据。注意模拟研究中假设每个密集追踪测量阶段内的数据符合平稳性假设，因此不需要进行去趋势处理(Zhou et al., 2021)。但实际研究仍需要对每个阶段采取去趋势的处理。因此，在前后测阶段分别加入测量时间点作为自变量进行去趋势化处理(Wang & Maxwell, 2015)。预实验结果如附表 11 所示。从表中可以看出，在当前的样本量水平下，反映均值差异干预效应的参数γ_{03}显著(γ_{03} = –0.155, CI = [–0.275, –0.031])，说明生态瞬时干预显著降低了社交媒体外貌焦虑水平。反映个体内变异性干预效应的参数ω_{05}显著(ω_{05} = –0.341, CI = [–0.586, –0.069])，说明干预显著降低了社交媒体外貌焦虑的个体内变异性。但是，反映自回归干预效应的参数γ_{41}不显著(γ_{41} = 0.023, CI = [–0.090, 0.134])，说明干预未能引起社交媒体外貌焦虑自回归的改变。

附表 11　社交媒体外貌焦虑的生态瞬时干预预实验参数估计结果

效应	参数	估计值	后验分布标准差	后验分布 95%可信区间宽度
固定效应	γ_{00}	**0.136**	0.046	[0.044, 0.224]
	γ_{10}	**–0.088**	0.042	[–0.172, –0.005]
	ω_{20}	**–0.252**	0.096	[–0.441, –0.065]
	γ_{03}	**–0.153**	0.062	[–0.275, –0.031]
	γ_{41}	0.023	0.057	[–0.090, 0.134]
	ω_{05}	**–0.341**	0.132	[–0.586, –0.069]
随机效应	τ^2_{00}	**0.805**	0.085	[0.660, 0.992]
	τ^2_{11}	**0.043**	0.009	[0.027, 0.062]
	τ^2_{22}	**0.927**	0.105	[0.756, 1.159]
	τ^2_{33}	**0.155**	0.020	[0.122, 0.199]
	τ^2_{44}	**0.041**	0.014	[0.016, 0.071]
	τ^2_{55}	**0.630**	0.091	[0.472, 0.835]

注：加粗的参数估计值是其后验分布 95%可信区间不包括 0 的情况，即显著的结果。

单臂设计数据生成和分析的 Mplus 语句(模型 1，以平衡设计情况下 N = 100, T = 200 为例)

```
MONTECARLO:
NAMES = y1 y2; !name variables to be simulated
NOBS = 10000; ! total number of measurement occasions
NREP = 500; ! number of replications
CSIZES = 100(100); !N(T)
```

[3] 反应时低于平均反应时 25%且 3 道测谎题均答错的被试将被视为不认真作答。

[4] 干预材料在相关文献材料的基础上进行编写和评定，结果证明自我同情语句与自我同情的相关程度显著大于中性语句与自我同情的相关程度。

```
NCSIZES = 1; ! number of unique cluster size
LAGGED = y1(1) y2(1); ! measurements at t-1
REPSAVE = ALL; ! save all generated datasets
SAVE=
D:/mplus/samplesize/s2/100-100-0.5/model-100-100.rep*.
dat;! location of generated datasets
RESULTS =
D:/mplus/samplesize/s2/100-100-0.5/output.sav;
BPARAMETERS
=D:/mplus/samplesize/s2/100-100-0.5/bayes.dat;
ANALYSIS:
TYPE = twolevel random;
ESTIMATOR = BAYES;
PROCESSORS = 2;
BITER = (10000);
BSEED = 5240;
MODEL POPULATION:
%within%
phi1|y1 on y1&1;
logv1|y1;
phi2|y2 on y2&1;
logv2|y2;
%between%
y1*0.04; [y1*0];
phi1*0.04; [phi1*0.2];
y2*0.01; [y2*0.112];
y2 on y1@1; ! treatment effect = 0
phi2 on phi1@1;
phi2*0.01; [phi2*0.112];
[logv1*0]; [logv2*0.112];
logv1*0.04; logv2*0.01;
logv2 on logv1@1;
MODEL:
%within%
phi1|y1 on y1&1;
logv1|y1;
phi2|y2 on y2&1;
logv2|y2;
%between%
y1*0.04; [y1*0];
phi1*0.04; [phi1*0.2];
y2*0.01; [y2*0.112];
y2 on y1@1;
phi2 on phi1@1;
phi2*0.01; [phi2*0.112];
[logv1*0]; [logv2*0.112];
logv1*0.04; logv2*0.01;
logv2 on logv1@1;
OUTPUT:
TECH1; !print parameter number and initial value
TECH8; !print optimization history
```

随机对照设计数据生成和分析的 Mplus 语句(模型 2，以平衡设计情况下 N=100, T=100 为例)

```
MONTECARLO:
NAMES = y1 y2 group; !name variables to be
simulated
NOBS = 10000; ! total number of measurement
occasions
NREP = 500; ! number of replications
CSIZES = 100(100); !N(T)
NCSIZES = 1; ! number of unique cluster size
LAGGED = y1(1) y2(1); ! measurements at t-1
BETWEEN = group; ! between-level variable
CUTPOINTS = group(0); !set group as a binary
variable with 50/50 split
REPSAVE = ALL; ! save all generated datasets
SAVE=
F:/LiuYue/EMI/RCT/100-100/model-100-100.rep*.dat;!
location of generated datasets
RESULTS =
F:/LiuYue/EMI/RCT/100-100/output.sav;
BPARAMETERS
=F:/LiuYue/EMI/RCT/100-100/bayes.dat;
ANALYSIS:
TYPE = twolevel random;
ESTIMATOR = BAYES;
PROCESSORS = 2;
BITER = (10000);
BSEED = 5240;
MODEL POPULATION:
%within%
phi1|y1 on y1&1;
logv1|y1;
phi2|y2 on y2&1;
logv2|y2;
%between%
group*1; [group*0];
```

```
y1*0.04; [y1*0];
phi1*0.04; [phi1*0.2];
y2*0.01; [y2*0];
y2 on y1@1 group*0.14; ! treatment effect = 0
phi2 on phi1@1 group*0.14;
phi2*0.01; [phi2*0];
[logv1*0]; [logv2*0];
logv1*0.04; logv2*0.01;
logv2 on logv1@1 group*0.14;
MODEL:
%within%
phi1|y1 on y1&1;
logv1|y1;
phi2|y2 on y2&1;
logv2|y2;
%between%
y1*0.04; [y1*0];
phi1*0.04; [phi1*0.2];
y2*0.01; [y2*0];
y2 on y1@1 group*0.14;
phi2 on phi1@1 group*0.14;
phi2*0.01; [phi2*0];
[logv1*0]; [logv2*0];
logv1*0.04; logv2*0.01;
logv2 on logv1@1 group*0.14;
OUTPUT:
TECH1; !print parameter number and initial value
TECH8; !print optimization history
```

随机对照设计下生成可信区间宽度等高线的 Mplus 语句

```
#数据导入
data <- read.csv("study2_gamma41_ci.csv")     #导入可信区间宽度数据
data <- as.data.frame(data)
power <- read.csv("study2_gamma41_power.csv") #导入检验力数据
names(power)[1] <- "x"
pt <- read.csv("study2_gamma41_pt.csv")
names(pt)[1] <- "x"

# 设置坐标轴和可信区间宽度等高线刻度
kd <- c(0.1, 0.15, 0.213, 0.5, 0.7, 0.9)
xbreak <- c(30, 60, 100, 150, 200, 300, 400)
ybreak <- c(20, 40, 80, 120, 160, 200)
label = xbreak
xlimit = c(0, 400)
ylimit = c(0, 200)

#绘制可信区间等高线图
plot1 <- ggplot() +
  theme_bw() +
  xlab("被试量") +# x 轴名称
  ylab("测量时间点数") +# y 轴名称
  stat_contour(
    data = data,
    aes(x = level2, y = level1, z = ci, colour = ..level..),#添加等高线
    breaks = kd,
    linewidth = 1.08
  ) +
  guides(color = guide_colorbar(#设置标题等
    title = "95%可信区间宽度",
    title.theme = element_text(size = 12),
    draw.ulim = TRUE,
    draw.llim = TRUE,
    reverse = TRUE
  )) +
  scale_color_gradientn(#设置等高线颜色
    colors = rev(c("#1822c7", "#c300a2", "#ff0073", "#ff714d", "#ffba43", "#f9f871")),
    breaks = kd
  ) +
  theme(#坐标轴样式
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    axis.line.x = element_line(size = 1),
    axis.line.y = element_line(size = 1),
    legend.text = element_text(size = 12),
    legend.key.height = unit(2, "cm")
  ) +
  scale_x_continuous(
    limits = xlimit, breaks = xbreak, labels = label, expand = c(0, 0)
  ) +
  scale_y_continuous(
    limits = ylimit, breaks = ybreak, expand = c(0,
```

```
0)
        ) +
        theme(
          axis.ticks.length.y = unit(-0.1, 'cm'),
          axis.ticks.length.x = unit(-0.1, 'cm')
        ) +
        geom_polygon(#添加阴影
          data = power, aes(x = x, y = y, group=1), fill =
"#edeed3", alpha = 0.3,color = "#edeed3", linewidth = 1.06
        ) +
        geom_point(#添加阴影轨迹点
          data = pt, aes(x = x, y = y), size = 1
        ) +
        geom_line(#添加阴影轨迹线
          data = pt, aes(x = x, y = y),
          color = "grey",
          linewidth = 1
        )
      plot1
```