收稿：制作时添加 修订：制作时添加 接受：制作时添加DOI：xxx/xxxx

研究论文

利用研究层面历史数据预测规划阶段临床试验的入组情况

Mengjia Yu | Sheng Zhong* | Yunzhao Xing | Li Wang*

艾伯维公司数据与统计科学部统计创新组，美国伊利诺伊州

通信作者信息
Li Wang，电子邮箱：wangleelee@gmail.com
Sheng Zhong，电子邮箱：zhongever@gmail.com

现地址
1 N Waukegan Rd, 北芝加哥，IL，60064

arXiv:2303.08065v1 [stat.AP] 14 Mar 2023

摘要

随着临床试验不断发展、需求日益增长，准确预测入组时间线对于战略决策和卓越的试验执行愈发重要。简单的方法假设入组率保持不变，并采用历史数据的平均值；传统统计方法则使用泊松-伽马模型，并假设中心启动率和受试者招募率不随时间变化。这两种方法都未考虑时间和地点等重要因素。我们提出了一种新颖的两阶段统计方法：采用准泊松回归估算受试者入组率，并采用泊松过程建模受试者入组和中心启动。输入的研究层面数据可公开获取，并可与用户组织的历史研究数据整合，从而前瞻性地预测入组时间线。与此前的研究相比，这一新框架简洁且准确。我们通过 7 项精选研究验证了所提出入组模型的性能，并将结果与其他框架进行了比较。

关键词：
准泊松回归、泊松过程、入组预测、自助法

1 引言

在临床开发中，尤其是在项目组合规划阶段，预测试验持续时间极为重要。这有助于高级管理层在充分了解信息的基础上做出决策，并初步了解任何新试验提案可能耗费的时间和成本。当前亟需一种先进的预测建模方法，能够为所有项目组合提供足够准确的入组预测。统计学文献中已提出多种定量方法。

一种简单的方法是按每个研究中心每月的患者（受试者）数计算，通常缩写为 psm，其定义为psm = 已入组患者或受试者数量/(研究中心数量 × 入组时间)。假设试验中的每个研究中心在整个期间都保持恒定的入组率，则规划中试验的预计入组时长只需用计划样本量/(拟设研究中心数量 × psm)计算，其中分母中的 psm 率可直接取自历史研究的平均率或用户以往的经验。这种方法简单且基于经验。它依赖于一个强假设，即所有研究中心的入组率在整个期间保持恒定，而现实中通常并非如此。鉴于这种简单方法的预测表现往往不尽如人意，需要采用更复杂的统计建模来弥补这一不足。

文献中提出了多种统计方法，用于建模和预测患者入组。论文[1,2]对患者入组模型进行了系统的文献综述。Bagiella 和 Heitjan[3] 提出使用齐次泊松过程对患者招募进行建模。Anisimov 和 Fedorov[4] 通过应用泊松-Gamma 混合模型改进了招募模型，以处理多个中心之间招募率的差异。正如后续研究所指出的，

这类方法[5,6,7], 主要是随机效应模型。其核心目标是捕捉不同中心之间入组率的异质性，而每个中心内部的入组模式则由速率服从 Gamma 分布的齐次泊松过程描述。Lan、Tang 和 Heitjan[8] 首次提出了时变速率函数，在考虑研究中心启动的同时，对招募过程中随时间衰减的趋势进行建模。Deng、Zhang 和 Long[9] 研究了一种采用非齐次泊松过程的贝叶斯方法，并在其框架中纳入了区域特异性入组。Zhang 和 Long[10,11] 采用非齐次泊松过程对患者入组建模，允许潜在的受试者累积入组率随时间变化。Wang 等人[12] 定义了终点成熟时间框架，并将这一概念与临床试验中的关键里程碑日期联系起来。他们提出了一种基于模拟的非齐次泊松过程，其入组率采用正态核，能够捕捉现实中的入组趋势起伏，并在模拟和真实研究的入组数据中提高了预测性能。

受 Wang 等人[12] 及其他人的研究启发，Zhong 等人[13] 提出了一种新颖的统计框架，该框架基于广义线性混合效应模型（GLMM），并通过贝叶斯分层建模框架运用非齐次泊松过程，在项目组合规划阶段预测试验持续时间。该方法利用专有数据中的站点级入组信息，基于一组预先选定的验证研究，更准确地预测试验持续时间。研究还表明，他们的建模与模拟方法能够恰当地校准数据变异性，并为不同名义水平的预测区间提供正确的覆盖率。

Zhong 等人[13] 面临的挑战之一，是站点级入组信息的可获得性及其规模。一些站点级数据具有专有性，可能无法公开获取；这些数据集中的研究数量也仅限于提供数据的几家申办方。为了让制药行业更广泛的用户受益于该预测算法，我们提出了一种新的先进统计建模算法，利用可从 ClinicalTrials.gov 获取的公开研究层面信息。与专有数据相比，这一数据来源免费，且包含大量研究。将 ClinicalTrials.gov 的研究层面信息与用户组织内部的历史数据结合起来，可对新研究中心及新适应症的受试者入组情况开展前瞻性评估。我们的方法学采用不同的齐次泊松过程来描述受试者入组和研究中心启动。我们提出使用准泊松回归和蒙特卡洛抽样来估计和模拟受试者入组，并使用线性（固定）方法和扰动方法来建模研究中心启动。

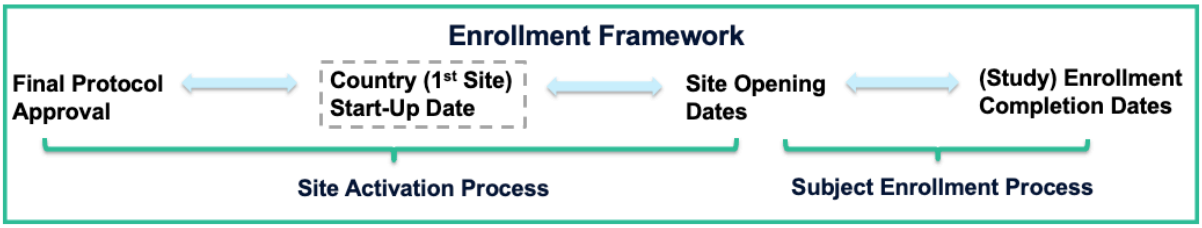
本文其余部分安排如下：第2节介绍输入数据和完整的预测工作流程。第3节详细阐述方法学。第4节根据实际案例研究给出预测性能。第5节进行讨论。

2 输入数据和预测工作流程

临床试验中的整个入组流程由多个步骤组成，是一个复杂的过程。以往研究通常侧重于受试者入组这一关键环节[5,6,7]。然而，受试者入组只是临床试验入组流程中的一个影响因素。在实际操作中，受试者入组之前的国家/研究中心准备步骤（例如签约、培训研究者、准备剂量等）也会对入组时间线产生重大影响。本文提出了一套基于研究层面历史试验数据的综合入组框架，涵盖研究中心启动和受试者入组流程。我们将根据历史数据的可用性，为入组框架的各个环节开发适当的模型，并进行参数估计和模拟。除非另有说明，如图 1 所示，研究中的整个入组过程是指从最终方案批准之日到受试者入组完成之日的两个连续阶段。

入组框架的第一部分是研究中心启动流程，定义为研究开始日期与各研究中心开放日期之间的时间。我们假设，临床试验中所有选定的研究中心及其所在国家/地区都采用相同的研究开始日期，即最终方案批准日期。此外，我们假设同一国家/地区内的研究中心彼此不可区分，并允许分别对各国/地区首个研究中心的启动日期进行建模。首个研究中心启动也称为国家启动，这是国家层面的流程，涉及国家批准以及准备活动，例如国家/地区监管批准和伦理委员会批准流程。在现实中，这一过程会受到多种因素的显著影响，包括但不限于申办方、国家/地区、治疗领域和试验阶段。因此，需要整合来自多个来源的数据，以预测国家启动和研究中心启动。从建模角度来看，数据可用性也是一个关键因素，因为最常见的公共数据库通常不提供最终方案批准日期和研究中心启动日期，

图 1入组框架的两个连续阶段



我们提出的模型将依托内部可用的数据集，开发研究中心启动预测流程。

入组框架的第二部分是受试者入组流程。这是一个国家特定流程，从国家启动日期开始，一直持续到整个临床试验的入组完成日期。同一国家的所有研究中心都会在国家启动日期之后开始招募患者。由于研究层面历史数据的限制，我们不考虑任何特定的研究中心启动模式。我们会根据模拟的国家启动模式调整历史试验的入组率，以避免偏差。

3 方法详情

3.1 总体描述

我们提出的框架旨在预测处于规划阶段的研究的入组时间线。已知各国家的研究中心数量后，只要能够对研究中心的启动时间以及各中心相应的受试者招募模式建模，就可以预测入组时长。在我们的框架中，我们假设各国家的研究中心数量是预先指定的，并且模型将在给定的治疗领域内构建。我们不假设一定能获得特定的招募标准或试验中观测数据。相反，内部和外部历史数据都会在每个步骤的参数估计中发挥作用。

我们的模型由两部分组成：各研究中心的受试者入组，以及各国家的中心启动。对于受试者入组模型，我们采用齐次泊松过程，并假设所有国家的所有研究中心的入组率均相同；该入组率可通过准泊松回归，利用历史研究数据进行估计。我们还考虑对受试者入组率进行蒙特卡洛抽样。对于中心启动模型，我们采用另一种齐次泊松过程，其入组率根据作者所在机构内部可用的历史数据进行估计和模拟。我们提出了一种简化的线性（固定）方法和一种扰动方法。接下来将介绍模型的详细内容，并在本节末总结建模流程。

3.2 受试者入组模型的泊松过程

设 $N_{ev}(u_1,u_2)$ 为时间点 u_1 和 u_2 之间入组的受试者人数。假设共有 n_{cntry} 个国家，其中第 i 个国家有 $n_{cntry\,个i}$ 研究中心开放招募受试者。设 $N_{ev,\ i,\ j}(u_1,u_2)$ 为第 i 个国家中第 j 个研究中心入组的受试者人数。于是，我们假设各研究中心的受试者入组过程相互独立，且均为泊松过程

$$N_{ev}(u_1,u_2)=\sum_{i=1}^{n_{cntry}}\sum_{j=1}^{n_{cntry,i}}N_{ev,i,j}(u_1,u_2),$$

$$N_{ev,i,j}(u_1,u_2)\overset{\text{independent}}{\sim}Poisson\left(\int_{u_1}^{u_2}\lambda_{ij}(v)dv\right),$$

其中，$\lambda_{ij}(=\lambda_i)$ 是同一县内所有研究中心相同的受试者入组率。

在我们的框架中，我们只采用各国相同的线性受试者入组率 $\lambda_i(v)=\mu$。临床试验中 μ 这一概念的一个常见示例是每个中心每月入组的患者数（psm）。给定中心启动（开放）时间

点 $u_{i,\ j},\ i = 1,\ \ldots,\ n_{cntry},\ j = 1,\ \ldots,\ n_{cntry,\ i}$，泊松过程中的率参数可简化为

$$\int_{u_1}^{u_2} \lambda_{ij}(v)dv = (u_2 \vee u_{i,j} - u_1 \vee u_{i,j})\mu,$$

wh这里，$a \vee b$ 是 a 和 b之间的最大值。因此，受试者入组模型变为

$$N_{ev}(u_1, u_2) \sim Poisson\left(\sum_{i=1}^{n_{cntry}} \sum_{j=1}^{n_{cntry,i}} \{u_2 \vee u_{i,j} - u_1 \vee u_{i,j}\}\mu\right).$$

这里，我们假设国家列表以及每个国家的研究中心数量通常由研究团队提供。如果这些信息尚待估计，我们将在第5节讨论国家-研究中心推荐算法。确定每个国家的研究中心数量后，只需知道 μ 和$u_{i,\ j}$即可生成受试者入组模式。

3.3 受试者入组率的准泊松估计 μ

由于我们假设计划中的研究是针对组织内的新药或新适应症，因此，与内部临床试验管理系统（CTMS）相比，Citeline（ClinicalTrials.gov）的全球数据是估计受试者累积入组率的更佳来源。为规避 Citeline 不包含研究中心层面信息这一局限，必须直接对受试者入组率 μ 建模。假设在此估计步骤中使用 $s = 1,\ \ldots,\ S$ 项历史研究，其中 S 项研究是根据预测研究的具体需求，按照特定标准从 Citeline 中筛选出来的。例如，筛选标准可包括研究阶段、治疗领域、适应症等。对于每项研究，将$T_{i\ ,\ j}^{(s)}$ 记为国家 i中研究中心 j 的入组时长，将 X_s 记为招募受试者人数，并将 d_s 记为研究中心数量和入组时长的综合效应偏移量。准泊松模型的设定为

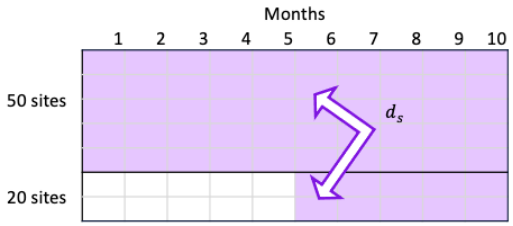
$$X_s \sim quasi - Poisson(\lambda_s), \quad E(X_s) = \phi\ var(X_s) \quad \text{and} \quad \log \lambda_s = \log d_s + \mu.$$

此处，我们通过离散参数 ϕ 允许存在过度离散效应，使得 $E(X_s) = \phi\ var(X_s)$ 并且

$$d_s = \sum_{i=1}^{n_{cntry}^{(s)}} \sum_{j=1}^{n_{cntry,i}^{(s)}} T_{i,j}^{(s)}.$$

偏移量 d_s 用于校准基于计数数据的速率 μ。在大多数情况下，广泛使用的 psm 只是用受试者总数除以研究中心总数和整个研究持续时间。 然而，由于运营考量及其他原因，各国家的启动速度可能不同。因此，进行任何调整前的 psm 可能较慢，导致预测与现实相差甚远。我们建议调整 d_s，其量纲应为“研究中心*月”，方法是结合各研究中心的开放时间；在数据可用时，可近似计算或精确计算。例如，假设一项研究的入组持续了10个月，其中国家1在第1天开放了50个研究中心，国家2在第6个月开放了20个研究中心。那么，如图2所示，该历史试验的 d_s 为600研究中心*月。理想情况下，如果每个研究中心的开放时间都有记录且可公开获取，最好根据这类细粒度信息计算 d_s。然而，有时甚至无法获知国家启动时间。在这种情况下，我们建议采用第3.4节中的方法来估计$T_{i\ ,\ j}^{(s)}$。

图2准泊松估计中受试者入组估计的偏置调整。



在实际应用中，当模型不包含直接对方差建模的参数时，自然会出现过度离散。广义线性模型中的一些常见情况包括基于多项分布的模型（如逻辑回归），其中方差是均值的函数；以及计数数据模型（如泊松回归），其中方差等于均值。为了在参数分布预测中更好地纳入变异性，我们采用蒙特卡罗模拟，根据模型估计值 $\hat{\mu}$ 及其对应的标准差 σ_μ抽取受试者入组率。例如，实践中可以从正态分布 $N(\hat{\mu}, \sigma_\mu^2)$ 中抽取μ_b，$b = 1, \ldots, B$。

备注*1*。 或者，随机 μ 的效应可引入泊松模型，以考虑过度离散效应，其中

$$X_s \sim Poisson(\lambda_s), \quad \log \lambda_s = \log d_s + \mu + \tau_s, \quad \tau_s \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\tau^2).$$

通过这一模型，不同研究的平均入组率可视为某一适应症类型或治疗领域类别总体层级效应的一次实现。与固定准泊松模型不同，在该模型中，我们只能估计具有共同方差的 S 项研究的受试者累积入组率均值；我们将方差分配给更广泛（但不可观测）的群体层级效应，从而将推断扩展到类似的试验集合。不过，随机效应模型通常需要适量的观测值。由于实际数据可用性有限，我们暂不考虑随机效应模型。适应症或治疗领域的随机效应可留待未来探索。

3.4 中心启动的泊松过程$u_{i,\ j}$ 以及线性（固定）投影

记$N_{site,\ i}$ 为国家 i 中已开放的研究中心数量，β_i 为相应的研究中心启动率。回顾一下，$u_{i,\ j}$ 表示国家 i中第 j个研究中心的启动时间。我们假设各研究中心不可区分，且其启动时间点服从以下泊松过程

$$N_{site,i}(u_2) - N_{site,i}(u_1) \sim Poisson\left((u_2 \vee t_i - u_1 \vee t_i)\beta_i\right),$$

其中，$t_i = u_{i,\ 1}$ 是国家 i中最早的研究中心启动时间，即国家 i的启动时间。将 w_k 记为研究 k的整体入组时长。因此，

$$u_{i,j} = \inf_u \{N_{site,i}(u) = j\}, \quad T_{i,j}^{(k)} = w_k - u_{i,j}.$$

在临床试验中，从方案批准到首个研究中心开放，准备期可能长达数月至一年甚至更久，所需准备工作包括但不限于建立数据库、开发试验主文件 (TMF)、培训主要研究者、制定受试者随机化时间表以及准备试验用药品套件。因此，为了提高准确性，我们必须将 t_i 与 $u_{i,\ j}$ 或 $T_i^{(k)}{}_{,\ j}$ 分别估计。由于数据可用性方面的原因，我们在此依赖内部临床试验管理系统数据，而非 Citeline 数据来提取锚定日期。我们的框架首先使用各国家内历史研究中心启动时间的中位数作为

$$\hat{t}_i = median_k\{t_i^{(k)} | t_i^{(k)} \text{ is the first site initiation time of Country } i \text{ for all available studies } k = 1, \ldots\}.$$

由于临床试验管理系统中每项研究的所有中心启动时间均可获取，因此也可以在汇总中找到首个研究中心启动与最后一个研究中心启动之间的时长。然后，其比率的中位数（即研究中心总数除以研究中心开放时长 ˆ）可用作 β_i的估计值。为了进一步简化泊松过程，可以使用线性（固定）投影来近似 $N_{site,\ i}(u)$ 和 $T_i^{(k)}{}_{,\ j}$，使任意相邻两次研究中心开放之间的时间相等。具体而言，只需估计 $\hat{u}_i$

同样，$\hat{t}_i + (j-1)\hat{\beta}_i$ 如果认为有必要，这个思路也可应用于上一节中校准$T_i^{(s)}{}_{,j}$ d_s。令$\hat{T}_i^{(s)}{}_{,j} = w_s - \hat{u}_{i,j}$，其中 w_s是研究 s，$s = 1, \ldots, S$的完整入组时长，而$\hat{u}_{i,j}$是国家 i中第 j个研究中心的启动时间。下标范围为 $i = 1, \ldots, n_{cntry}, j = 1, \ldots, n_{cntry,i}^{(s)}$，其中$n_{cntry,i}^{(s)}$是历史研究 s在国家 i 启动的研究中心总数。在实际应用中，通常可以获得 w_s，但Citeline提供的$n_{cntry,i}^{(s)}$有时可能不可靠。因此，还可以通过令$\hat{T}_i^{(s)}{}_{,j} = w_s$来进一步近似 d_s，这样 d_s 中的求和就可以简化为试验持续时间与研究中心总数的乘积。在我们的数值研究中，我们直接采用了来自内部临床试验管理系统数据的估计值。

3.5 中心开放速率 β_i 和研究中心启动的国家层面扰动 t_i

在模型开发中，当数据与模型一致时，人们通常认为，数据点越多，所描绘的情况就越准确、越有说服力。然而，在实际应用预测模型时，狭窄的预测窗口（预测区间）可能会限制其对团队开展研究准备或预算计算的指导作用。另一个批评是，未来

预测结果可能会悄然偏离历史记录。因此，我们建议基于历史国家级数据进行参数扰动，引入一定的随机性，以避免这种虚假的精确性问题。

例如，我们可以对国家 i 的研究中心启动时间（或中心开放速率）进行自助抽样，将其作为 t_i（或 β_i）的估计值，而不是使用固定的中位数。我们还可以对每个国家的 t_i 和 β_i 进行配对抽样，以保持两者之间通常为正的相关性。事实上，在临床试验中，研究中心准备工作（如签约和培训）总是在方案批准前很久就开始了。某个研究中心（在某个国家）的启动延迟可能不会严重影响其他研究中心尽快做好运营准备，因此，β_i 通常会呈现较快的开放模式，同时伴随较长的国家启动时间 t_i。此外，如果试验推迟研究中心启动，研究团队很可能会催促后续研究中心开放并招募患者，以赶上时间表。因此，对研究中心启动时间和中心开放速率进行配对自助抽样更为理想。

3.6 建模流程总结

算法1为我们的研究级入组预测模型提供了伪代码。

算法1 研究级入组模型算法

需要： $u_1 \leftarrow 0$ ▷ 例如，时间零点可以是方案批准日期。
需要： 治疗领域、国家列表等，用于筛选历史数据
确保： 预期的研究中心数量 $n_{cntry,i}$ 针对 $i = 1, \ldots, n_{cntry}$ 以及入组受试者 n_{ev} 已给出。
对于 $b = 1, b++, b \leq B$ **执行**
中心启动：
估计 $\hat{t}_i$ 对于 $i = 1, \ldots, n_{cntry}$
估计 $\hat{\beta}_i$ 针对 $i = 1, \ldots, n_{cntry}$
$\hat{u}_{i,j} \leftarrow \hat{t}_i + (j-1)\hat{\beta}_i$ 和 $T_i^{(s)}{}_{,j} = w_s - u_{i,j}$ 针对 $j = 1, \ldots, n_{cntry,i}$
获取历史研究的偏移量 $\hat{d}_s$ 用于 $s = 1, \ldots, S$
受试者入组：
拟合准泊松模型并进行估计 $\hat{\lambda}_s$ 面向 $s = 1, \ldots, S$
分别抽取样本 $N_{ev,i,j}(u_1, u_2)$ 并求和
结束循环

4 性能评估

4.1 总体考虑与数据描述

我们将把研究层面预测模型应用于所在机构近期完成的 7 项研究，并将预测结果与其实际入组时长进行比较；实际入组时长是指从末例受试者首次给药（LSFD）到方案批准之间的时间段。

用于估算中心启动情况的数据为作者所在机构内部的临床试验管理系统（CTMS）数据。在受试者入组建模过程中，我们筛选出与待预测研究处于相同治疗领域（肿瘤学、神经科学、免疫学和普通医学）且研究阶段相同（II 期和/或 III 期）的干预性研究。用于受试者入组建模的数据包括内部数据以及 Citeline 上可获取的外部数据。相关信息包括治疗领域、研究阶段、研究持续时间、受试者总数、研究中心总数、疾病适应症和患者群体。所选研究必须为已完成的干预性 II 期和/或 III 期研究，并且治疗领域、疾病适应症和患者群体均与待测试研究相同。用于性能测试的 7 项候选研究由独立的数据管理团队筛选和整理，以确保数据的客观性和完整性。出于保密原因，研究编号已隐去。

我们还将所提出建模框架的性能与先前内部开发的建模流程进行比较，该流程基于经典的泊松-伽马模型以及一种新的研究中心级建模器[13]。

4.2 模型性能

4.2.1 采用线性（固定）方法估计研究中心开放

我们首先应用模型，并按照第 3.4 节所述，使用 t_i 和 β_i 的线性（固定）估计。表1中，我们报告实际入组时长，以及预测值和相应的预测误差。我们还提供 95%预测区间（PI），并列出在 95%预测区间内以及在 +/- 1 个月、 +/- 2 个月和 +/- 3 个月范围内的覆盖状态。为进一步探究建模有效性的来源，我们还列出实际 FSFD（首位受试者首次给药）和预测 FSFD，以及相应的 95%预测区间，以反映研究中心启动 t_i方面的预测准确性。

从入组预测结果可以看出，与研究1相比，线性方法对研究2至7的估计更准确。95%预测区间涵盖了4项研究，但这些预测区间的范围也较宽，可能缺乏实用性。因此，我们还应关注固定窗口覆盖率。总体而言，3个月覆盖率表明我们的入组预测效果良好。比较实际 FSFD（首位受试者首次给药）与预测 FSFD 后发现，线性（固定）方法的预测通常较为激进，这表明中心启动部分存在负偏差。

研究	实际值 入组人数-ment (mo)	预测值 入组人数-ment (mo)	预测误差 (mo)	95%预测区间 (mo)	在95%预测区间	在+/-1月	在+/-2月	在.....范围内 +/-3月	实际值 FSFD (mo)	预测值 FSFD (mo)	95%预测区间 (mo)
1	26.4	19.3	-7.0	(15.5,24.0)	否	NO	NO	NO	5.3	4.6	(4.1,6.2)
2	17.5	18.9	1.4	(15.8,22.4)	是	NO	YES	YES	7.9	4.2	(3.1,5.3)
3	11.3	9.2	-2.1	(8.1,10.3)	NO	NO	NO	YES	6.3	3.1	(3.0,3.8)
4	10.0	8.7	-1.3	(7.7,9.9)	NO	NO	YES	YES	7.0	3.2	(3.0,4.0)
5	21.1	20.1	-1.0	(16.0,26.5)	是	NO	YES	YES	8.4	4.1	(3.1,5.2)
6	16.1	19.0	2.9	(14.9,24.1)	是	NO	NO	YES	6.0	5.4	(4.1,6.7)
7	10.8	11.2	0.3	(8.0,18.1)	YES	YES	YES	YES	5.4	4.5	(4.0,5.6)

表1 使用线性（固定）方法的研究层面预测性能。

4.2.2 基于扰动的研究中心开放方法

接下来，我们使用第3.5节介绍的基于扰动的模拟来应用该模型。表2中给出了对7项研究进行预测的类似展示。

入组预测显示，与研究1相比，我们的基于扰动的方法在研究2至7中总体表现良好。95%预测区间覆盖率和2个月覆盖率表明，我们的预测能够捕捉数据的变异性，同时提供准确的估计。将实际 FSFD 与预测 FSFD 进行比较时，基于扰动的预测仍然偏激进，但相应的95%预测区间比线性（固定）方法更宽。

表3从汇总层面对两种方法进行了比较。扰动方法的95%预测区间中位数长度更宽，且其中位数预测误差更小，说明其预测更准确。这表明，扰动方法能够捕捉研究中心启动时间与中心开放速率之间的正相关性，因而有所改进。此外，除3个月覆盖率（研究6）外，扰动方法的覆盖率均优于固定方法。我们得出结论，两种方法表现相当，但扰动模型在入组方面能提供略好的点预测和更宽的预测区间。

在实际应用中，稍窄一些、约为 +/-3 个月的预测区间，或许更适合为运营工作提供指导。可以降低预测区间计算的置信水平。例如，扰动估计方法得出的70%和80%预测区间的中位数长度分别为5.3和6.9，而7项研究中有5项被其各自的70%和80%预测区间覆盖。因此，从应用角度来看，预测区间覆盖率与预测区间长度之间的权衡需要针对具体情况进行评估。

研究	实际值入组-ment (mo)	预测值入组-ment (mo)	预测误差 (mo)	95%预测区间 (mo)	在95%预测区间	在+/-1月	在其中+/-2月	内部+/-3月	实际值FSFD (mo)	预测值FSFD (mo)	95%预测区间 (mo)
1	26.4	18.5	-7.9	(14.7,24.9)	否	NO	NO	NO	5.3	3.3	(1.7,6.9)
2	17.5	18.8	1.3	(15.3,23.1)	是	NO	YES	YES	7.9	4.1	(1.3,7.3)
3	11.3	9.7	-1.5	(6.7,15.6)	YES	NO	YES	YES	6.3	3.3	(1.1,9.5)
4	10.0	9.4	-0.6	(6.1,15.9)	YES	YES	YES	YES	7.0	4.1	(1.2,10.5)
5	21.1	20.1	-1.1	(16.0,26.5)	是	NO	YES	YES	8.4	4.2	(1.2,7.2)
6	16.1	19.3	3.2	(15.4,25.2)	是	NO	NO	NO	6.0	5.5	(4.1,9.4)
7	10.8	10.3	-0.5	(6.0,19.7)	YES	YES	YES	YES	5.4	3.6	(1.4,7.7)

表2 使用扰动方法的研究层面预测性能。

	95%预测区间长度 中位数（月）	预测误差 中位数（月）	覆盖率 95%预测区间	+/-1 个月	+/-2 月	+/-3 月
固定估计	8.46	-1.02	57%	14%	57%	86%
扰动估计	9.78	-0.60	86%	29%	71%	71%

表3 两种提出方法的预测性能比较。

4.3 与现有模型的比较

我们将提出的扰动方法与Zhong等人[13]所述的站点级模型和传统模型进行比较。为确保公平，我们将重点关注相同的7项研究，即Zhong等人[13]中的研究6、13、14、15、16、24和25。表4列出了95%预测区间长度和绝对预测误差的均值及中位数（单位为月，mo），随后列出95%预测区间、 +/-1个月、 +/-2个月和 +/-3个月的覆盖率。站点级模型和传统模型的详细预测结果分别见表5和表6。

	95%预测区间长度（月）		绝对预测误差（月）		覆盖率			
	中位数	mean	中位数	mean	95%预测区间	+/-1 月	+/-2 月	+/-3 月
扰动方法	9.8	10.1	1.3	2.3	86%	29%	71%	71%
站点级模型	17.6	Inf	1.1	4.3	100%	29%	57%	71%
传统模型	1.6	2.4	4	5.3	29%	29%	29%	43%

表4 Zhong等人[13]提出的研究层级模型、站点级模型与传统模型之间的比较。

从表4中，我们可以得出以下观察结果。传统模型的95%预测区间长度最短，但其预测偏差（以绝对预测误差衡量）最大，且在全部4项标准下的覆盖率最低。因此，该模型往往会产生有偏预测，无法捕捉数据的变异性。站点级模型的95%预测区间最宽，因此相应的覆盖率达到100%。然而，正如作者所提到的，站点级模型考虑了许多随机性来源，这些随机性来源累积于由国家启动、研究中心启动和用于受试者入组的混合效应泊松回归构成的三阶段框架，以及招募的蒙特卡洛模拟。因此，站点级模型的某次特定模拟可能会偶然陷入停滞状态，无法在足够长的时间内（本例中为2000天）招募足够的样本。这导致研究6的预测区间上限为无穷大。但从绝对预测误差中位数和 1~3 月覆盖率来看，站点级模型总体上能够提供良好的点估计。我们提出的扰动方法模型可以生成适中的95%预测区间，其绝对预测误差与站点级模型相当。我们的 2~3 月覆盖率是最高的

在所有三种方法中。因此，我们提出的方法既保留了预测精度，也在预测区间长度与预测区间覆盖率之间实现了良好的平衡。

注释 *2.* 我们希望从模型本身出发，深入比较站点级框架与研究级框架。首先需要强调的是，应考虑数据来源的差异。实际上，在实践中很难获取站点级数据，因此站点级模型所依据的历史研究数量通常远少于研究层级模型。因此，站点级模型的 95%预测区间较宽。其次，从方法学角度来看，站点级模型更复杂。站点级框架由三个环节组成（即国家启动、中心启动和受试者入组），能够适用于广义上的入组过程。由于实践中缺乏包含站点级数据的研究，其数值表现可能尚未达到极限。因此，从业者在确定使用哪种模型之前，应评估数据可用性。

研究	实际值入组 (mo)	预测值入组 (mo)	预测误差（个月）	95%预测区间（个月）	在95%预测区间	在...之内+/-1月	在.....之内+/-2月	在其中+/-3月
1	26.4	20.5	-5.9	(13.7,31.3)	YES	NO	NO	NO
2	17.5	20.2	2.7	(13.5,35.4)	YES	NO	NO	YES
3	11.3	10.1	-1.1	(6.4,17.3)	YES	NO	YES	YES
4	10.0	10.0	-0.0	(6.9,17.3)	YES	YES	YES	YES
5	21.1	22.1	1.0	(15.7,36.7)	YES	NO	YES	YES
6	16.1	35.0	18.9	(16.2,Inf)	YES	NO	NO	NO
7	10.8	11.6	0.8	(8.11,17.7)	YES	YES	YES	YES

表5 站点级模型 Zhong等人[13]的表现。

研究	实际值入组 (mo)	预测值入组 (mo)	预测误差（月）	95%预测区间（月）	在95%预测区间	范围内+/-1月	在.....之内+/-2月	在.....内+/-3月
1	26.4	17.7	-8.7	(16.9,18.4)	NO	NO	NO	NO
2	17.5	13.5	-4.0	(12.7,14.2)	NO	NO	NO	NO
3	11.3	9.1	-2.2	(8.5,9.9)	NO	NO	NO	YES
4	10.0	9.6	-0.4	(8.8,10.4)	YES	YES	YES	YES
5	21.1	15.3	-5.8	(14.5,16.1)	NO	NO	NO	NO
6	16.1	31.1	15.0	(27.4,35.0)	NO	NO	NO	NO
7	10.8	11.7	0.9	(10.8,12.7)	YES	YES	YES	YES

表6 Zhong等人[13]的传统模型性能。

5 讨论

针对在规划阶段早期预测研究持续时间的问题，我们提出的研究层级预测模型简洁且准确，能够通过受试者招募和研究中心启动这两个部分来刻画整个入组过程。通过数值分析，我们的模型在预测区间长度与预测误差之间实现了平衡，这本质上体现了方差与偏差之间的权衡。

在确定每个国家的研究中心数量时，我们的框架假设这些数量在一开始就已确定。然而，在研究准备的早期阶段，通常只有一份计划开展试验的国家名单和待启动的研究中心总数

。因此，需要从历史数据中获取更多见解，以指导国家启动和研究中心启动。在我们的性能评估中，我们使用了给定国家名单中的研究中心数量，并根据名单中所有国家的模型参数中位数估计值，估算其余研究中心的数量。同样，从业者可以利用历史数据，依据例如每项研究中已开放研究中心的最大数量、中心开放速率和国家启动天数等指标，获得关于各国研究中心数量的建议。

关于研究中心启动的预测，我们在第4.2节的性能评估中发现，FSFD似乎存在负偏差，尽管仅基于7项研究进行验证，尚无法得出确切结论。尽管如此，研究持续时间的总体预测通常较为准确。这一现象表明，研究中心激活建模仍有改进空间。一种可能的补救方法是，像Zhong等人那样，除研究中心启动外，另行建立国家启动模型[13]，以更好地建模t_i。另一种方法是针对$u_{i,\ j}$考虑使用零膨胀泊松模型，在该模型中引入一个额外的潜在过程，以确定计数是否为零。随后，该潜在过程会提高观测值为零的概率，而较大的率参数则会起到平衡作用，从而维持整体建模效果。由于这超出了本项目的范围，我们将零膨胀模型的训练留待未来进行。

在受试者入组率参数方面，我们的模型将 μ设定为各国家通用的参数。一个自然的变化是结合国家因素来估计该率。然而，正如准泊松模型的设定中所提到的，历史数据的可用性并不支持纳入过多因素（如随机效应和国家效应）的整合结构。因此，我们在此谨作上述修改建议，供参考。

披露

本稿件由艾伯维赞助。艾伯维参与了研究设计、研究实施、数据解读、撰写和审阅，并批准了稿件内容。所有作者均为艾伯维公司员工，并可能持有艾伯维股票。

数据可用性声明

作者选择不共享数据。

参考文献

1. Anisimov VV. 对 DF Heitjan、Z Ge 和 GS Ying 所著论文《临床试验入组和事件计数的实时预测：综述》的讨论。*Contemporary clinical trials* 2016年；46：7–10。2. Heitjan DF、Ge Z、Ying sG. 临床试验入组和事件计数的实时预测：综述。*ContemporaryClinical Trials* 2015年；45：26 - 33。十周年特别刊，doi：https://doi.org/10.1016/j.cct.2015.07.010 3. Bagiella E、Heitjan DF. 随机临床试验分析时间的预测。*Statistics inMedicine* 2001年；20(14)：2055- 2063。doi：https://doi.org/10.1002/sim.843 4. Anisimov VV、Fedorov VV. 多中心试验招募的建模、预测与自适应调整。*Statisticsin Medicine* 2007年；26(27)：4958-4975。doi：https://doi.org/10.1002/sim.2956 5. Anisimov VV. 临床试验运营预测建模的现代分析技术。载于：Springer。2020年（第361–408页）。6. Anisimov VV. 临床试验的统计建模（招募与随机化）。*Communications inStatistics-TheoryandMethods*2011年；40(19-20)：3684–3699。7. Anisimov VV、Downing D、Fedorov VV. 多中心试验中的招募：预测与调整。载于：Springer。2007年（第1–8页）。8. Lan Y、Tang G、Heitjan DF. 临床试验招募的统计建模与预测。*Statistics inMedicine* 2019年；38(6)：945-955。doi：https://doi.org/10.1002/sim.8036 9. Deng Y、Zhang X、Long Q. 多区域临床试验受试者累积入组的贝叶斯建模与预测。*Statistical methodsin medical research* 2017年；26(2)：752–765。10. Zhang X、Long Q. 临床试验受试者累积入组的随机建模与预测。*Statistics in Medicine* 2010年；29(6)：649–658。11. Zhang X、Long Q. 临床试验受试者累积入组与事件时间的联合监测和预测。*Biometrical journal* 2012年；54(6)：735–749。12. Wang L、Liu Y、Chen X、Pulkstenis E. 临床试验终点成熟时间的实时监测与预测。*Statistics in Medicine*2022年；41(18)：3596–3611。13. Zhong S、Xing Y、Yu M、Wang L. 基于中心级历史数据的项目组合规划阶段临床试验入组预测。审稿中2023年。

