收稿：添加于制作阶段 修订：添加于制作阶段 接受：添加于制作阶段

DOI：xxx/xxxx

研究论文

项目组合规划阶段基于研究中心历史数据的临床试验入组预测

Sheng Zhong | Yunzhao Xing | Mengjia Yu | Li Wang

艾伯维公司，美国伊利诺伊州，数据与统计科学部统计创新组

通讯
Li Wang，邮箱：wangleelee@gmail.com；
Sheng Zhong，邮箱：zhongever@gmail.com

现地址
伊利诺伊州北芝加哥市 N Waukegan Rd 1号，邮编 60064

arXiv:2301.01351v1 [stat.AP] 3 Jan 2023

摘要

在规划阶段准确预测临床试验入组时间线，对于企业战略规划和临床试验运营卓越都至关重要。末例受试者首次给药日期等关键里程碑的预测结果可为战略决策提供依据，而详细的预测洞见（例如某个国家每月的入组受试者人数中位数）则有助于规划临床试验运营活动并促进执行卓越。一种朴素方法通常根据历史数据计算平均入组率，并基于该平均入组率的线性趋势生成不准确的预测结果。传统统计方法采用简单的泊松-伽马模型，假设研究中心启动率不随时间变化，因此可能无法捕捉潜在的非线性模式（例如研究中心启动模式的上升和下降）。我们提出了一种新颖的统计方法，该方法基于广义线性混合效应模型，并通过贝叶斯框架使用非齐次泊松过程，以系统化的方式依次对国家启动、中心启动和受试者入组进行建模。我们基于从四个治疗领域中预先选定的25项研究，验证了所提出入组建模框架的性能。与传统统计方法相比，我们的建模框架在预测准确度方面取得了显著提升。此外，我们还表明，我们的建模与模拟方法能够适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。最后，我们展示了如何应用该方法生成随时间变化的预测入组曲线，并叠加置信带。

关键词：
非齐次泊松过程、贝叶斯层次模型、广义线性混合效应模型、入组、投资组合规划

1 引言

在临床开发的投资组合规划阶段，预测试验入组时长及其相关成本是高级管理层在决定是否为试验提供资金前必须考虑的关键可行性因素。因此，任何制药公司都亟需一种预测建模算法，能够为所有投资组合提供足够准确的入组预测，以协助管理层做出决策。目前实践中采用的最简单、最直接的方法基于

平均入组率这一概念，即一项研究中每个研究中心每月入组受试者数量的平均值，通常缩写为 *psm*, $psm = \frac{\text{入组受试者人数}}{\text{研究中心数量} \times \text{入组时长}}$；该方法还假设试验中的每个研究中心在整个期间内都保持恒定的入组率。计划中的试验的预计入组时长只需用$\frac{\text{计划样本量}}{\text{拟设研究中心数量} \times psm}$计算，其中分母中的 *psm* 率可以直接取自历史研究的平均入组率、根据专家知识指定，或通过混合方法获得。这是一种快速简便的方法，可根据新试验的样本量和拟设研究中心总数来预测入组时长，无须深入的统计思考或理论支撑。从方法学角度看，这种方法纯属经验性方法，无法量化不确定性（如置信区间），并且依赖于所有研究中心的入组率始终恒定这一强假设，而现实中往往并非如此。这通常是制药公司内部采用的第一种工具，主要由接受过较少正规统计学训练的团队实施。由于这种简单方法的预测表现不尽如人意，企业通常会转而咨询更侧重统计建模的团队。因此，开发一种更系统、更准确的统计预测建模方法，既有必要，也面临着机遇与挑战。

文献中已提出多种用于建模和预测患者入组的统计方法。Heitjan 等人的综述论文对现有患者入组模型进行了系统而出色的总结。Bagiella 和 Heitjan 使用齐次泊松过程对招募进行建模。Anisimov 和 Fedorov 通过采用泊松-伽马混合模型，改进了招募模型，以处理多个中心之间招募率的差异。这类方法基本上采用随机效应模型，旨在捕捉不同中心之间入组率的异质性，同时用速率服从伽马分布的齐次泊松过程描述各中心内部的入组模式。Lan 等人采用时变速率函数，在考虑中心启动的同时对招募中的时间衰减趋势进行建模。Deng 等人研究了基于贝叶斯方法的入组建模，并在其框架中使用非齐次泊松过程来考虑区域特异性入组。Zhang 和 Long 使用非齐次泊松过程对患者入组进行建模，并允许潜在入组率随时间变化。Wang 等人定义了终点成熟时间框架，并将其与临床试验的关键里程碑日期相关联。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，可以捕捉实际入组趋势的起伏，并在模拟和真实研究的入组数据中提高了预测性能。上述近期研究促使我们通过贝叶斯框架应用非齐次泊松过程对中心启动流程进行建模，并进行了一些技术调整，以便将多个历史研究作为输入数据。

然而，上述建模框架主要用于试验执行阶段的试验监测，此时患者已实际入组试验，目标是利用截至某一时间点累积的试验中数据，预测或重新预测此后未来的入组情况。目前，关于项目组合规划阶段入组预测的已发表文献还不多。为填补这一空白，我们提出了一种新颖的统计框架，基于广义线性混合效应模型（GLMM），并通过贝叶斯分层框架运用非齐次泊松过程，利用历史中心层面的入组相关数据，以系统化的方式依次对国家启动、中心启动和受试者入组进行建模和预测。我们从四个治疗领域中随机选取了25项已完成的研究，以验证所提出建模框架的性能。具体而言，我们将该框架的预测准确度与本公司先前开发的一套预测系统进行比较；该系统基于传统统计方法，通过拟合简单的概率分布进行预测。此外，我们表明，我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。最后，我们展示了如何运用该方法生成随时间变化的预测入组曲线，并叠加置信带。

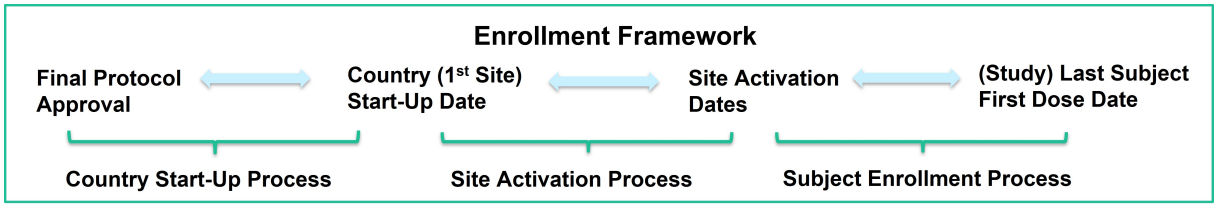
2 输入数据与预测工作流程

2.1 入组框架

临床试验中的整个入组流程并不仅仅是一个简单的受试者入组步骤。相反，这是一个包含多个步骤的复杂过程。为促进后续的统计模型开发，首先需要明确一个全面的入组框架，清晰界定整个入组流程中的每个步骤。在文献中，人们主要关注受试者入组过程[5,6,7]。然而在实际应用中，受试者入组之前的若干重要运营步骤（如国家启动和中心启动）所需的时间会对试验持续时间产生至关重要的影响。在本文中，我们对试验持续时间进行了拆解，并建立了一个全面的入组框架，将临床试验入组的所有

重要运营步骤纳入其中。针对整个入组框架中的每个阶段，我们均开发了先进的统计模型（具体取决于历史数据的可用性）。该入组框架定义为从最终方案批准日期到研究最后一名受试者首次给药日期（LSFD）的整个时间段，包含三个依次衔接的阶段，如图1所示。

图1入组框架的三个连续阶段



入组框架的第一阶段是国家启动流程。国家启动时间定义为研究启动日期至国家启动日期之间的时间。在我们的考虑中，临床试验中所有选定国家共用同一个研究启动日期，即最终方案批准日期。另一方面，每个国家都有自己的国家启动日期；由于我们的建模历史数据中缺少明确的国家启动日期，因此在本例中将其定义为该国首个研究中心启动日期。国家启动是国家层面的流程，涉及国家审批和准备活动，例如国家/地区监管批准和机构审查委员会审批流程。国家启动可能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。就建模而言，数据可用性也是需要考虑的重要因素，因为最终方案批准日期和研究中心启动日期通常无法从最常见的公共数据库中获取。因此，需要整合来自多个数据来源的专门数据，以估算该国的启动时间。

国家启动后紧接着，入组框架的第二个部分是中心启动流程。可以将其视为一个随机过程：从研究中心所在国家的启动日期开始，研究中心陆续开放。中心启动流程描述了临床试验中若干中心层面准备活动，例如文件移交、中心培训和药物配送。它可以反映研究中心的相对灵活性，通常因不同治疗领域、适应症、国家和地区而异。定义中的中心启动日期，是指研究中心准备就绪并开放参与临床试验患者招募的日期。

入组框架的最后一个部分是受试者入组过程。这是一个中心特定流程，从中心启动日期开始，到整个临床试验的最后一名受试者入组日期为止。如上所述，研究中心在中心启动日期后开始招募患者。我们认为，每个已启动研究中心的患者入组流程将持续到整个试验最后一名患者入组的时间。这意味着，整个临床试验的 LSFD 日期被用作所有研究中心患者入组的结束日期。尽管当前数据集中也有每个研究中心最后一名受试者的入组日期，但不建议将其用作患者入组流程的结束日期，因为研究中心一旦启动，就有能力在整个试验入组结束前持续招募患者。在某个特定研究中心最后一名患者入组与整个试验患者入组结束之间的时间段内，该研究中心仍处于患者入组流程中（尽管已无法再招募患者）。如果不考虑这一时间段，就会给各研究中心的患者入组时长带来额外偏差。

2.2 历史数据

我们利用两大历史数据来源对入组框架中不同阶段的持续时间进行建模：来自内部临床试验管理系统（CTMS）的中心级数据，以及来自 IQVIA Inc 专有数据查询系统（DQS）的中心级数据，后者包含多个申办方的入组数据。尽管每种数据来源都有各自的局限性，但将这两种数据来源结合起来，可提供各阶段最全面的中心级入组信息，从而能够对所提出的入组框架进行建模。

在 CTMS 数据中，可以获取每项临床试验详细的研究中心级时间线。这些数据提供了入组框架所需的锚定日期和流程信息，尤其是每项试验最终的方案批准日期。然而，这些数据

仅限于艾伯维开展的历史临床试验。若某一适应症或国家/研究中心没有艾伯维既往经验，仅凭这些数据来规划相关情景将颇具挑战。

另一方面，在 DQS 系统中，包括艾伯维在内的 17 家大型制药公司已签约彼此共享其临床试验的运营数据。来自不同申办方的数据会被传输并整合到一个数据库中。尽管各申办方内部系统中的字段名称和定义可能有所不同，但在传输到 DQS 之前，这些字段已得到良好协调，并映射到一套标准的字段名称和定义中。此外，DQS 采用了一种精心设计的算法，用于在不同临床试验和申办方之间识别临床研究中心，并为其分配唯一标识符。这使得不同临床试验和申办方中同一研究中心的历史信息能够叠加呈现。

在所提出的入组框架中，艾伯维 CTMS 数据用于对国家启动时间进行建模。DQS 数据包含艾伯维内部和外部研究的数据，用于对中心启动流程和患者入组流程进行建模。

3 方法学详情

3.1 一般考虑

我们的首要目标是在项目组合规划阶段预测一项研究的入组时间线，此时关于计划研究的信息非常有限。这意味着无法获得任何详细的招募规划信息（例如各国计划设置的研究中心数量），而且该研究的实际试验期间入组数据（例如截至某个截止日期实际已启动的研究中心数量）尚未观测到；试验期间入组数据可包括国家启动、中心启动和受试者入组等方面的数据。为进行入组预测，会提供有关计划研究的非常基本的信息，包括计划入组的受试者总数、计划设置的研究中心总数、患者人群、疾病适应症以及研究的治疗领域。

由于在规划阶段无法获得试验期间入组数据，为实现我们的首要目标，有必要利用内部或外部历史研究中的入组相关信息，对入组框架中定义的潜在入组相关过程进行建模。针对每个细分群体估计模型后，我们采用基于蒙特卡洛模拟的方法，模拟并预测计划研究未来的入组活动。

作为次要目标，除了总体受试者入组时间表外，预测受试者来自哪些具体国家，以及各国启动的研究中心数量，通常也备受临床运营团队关注。这些国家层面的详细见解有助于规划临床试验运营活动，并促进执行卓越。

国家层面预测性见解的需求对统计建模的启示是，只需对国家特异性效应进行建模即可，尽管我们的历史研究入组数据会通过唯一的研究中心标识符区分不同研究中的研究中心。因此，对于某项研究，在一个国家内观察到的研究中心将被视为无法相互区分，且是从该国可用研究中心中随机抽取的。采用这种国家层面的建模考量，可避免对研究中心特异性效应进行不必要的技术性复杂建模；而这对于实现我们的首要或次要目标都并非必要。

3.2 国家启动时间模型

假设我们观察到来自 S 项历史研究的入组相关数据，这些历史研究通常根据其治疗领域、疾病适应症和患者亚群（例如儿科患者）是否与当前计划研究相同来选取。其他研究层面特征，例如研究阶段（如 2b期）或研究启动年份，可用于选取近期研究，这些研究可能具有相似的招募特征。

对于每项历史研究 $j \in \{1, ..., S\}$，均观察到有 N_j 个国家启动，并且这些国家的国家启动日期均可获得。不同历史研究中已启动的国家构成一个包含 C 个不同国家的候选池，记为 $\{1, ..., C\}$。观测到的历史研究数据来自 $N = \sum_{j \in S} N_j$ 个研究-国家组合的列表。对于第 i 个研究-国家组合，研究索引映射 $j[i] \in \{1, ..., S\}$ 给出研究索引，国家索引映射 $k[i] \in \{1, ..., C\}$ 给出国家索引。需要注意的是，采用这一记法是因为不同历史研究为开展受试者入组而启动的国家集合各不相同（但彼此可能存在重叠）。对于一项历史临床试验 j，令 $t_{0,\,j}$ 表示方案批准日期，该日期被视为与临床试验招募相关活动的起始点。我们将研究中某个国家的国家启动时间定义为研究方案批准日期与国家启动日期之间的持续时间。在当前情况下，国家启动日期取该国的首个研究中心启动日期，因为我们的历史

数据不包含国家启动的具体日期（即其首个研究中心启动日期之前的日期）。对于收集了此类国家启动具体日期的情形，我们将在最后一节讨论其他建模选择。令 t_i 表示第 i 个研究-国家组合观测到的国家启动日期，令 $u_i = (t_i - t_{0,j[i]})/m$ 表示相应的观测到的国家启动时间，其中 m 是提供所需时间单位的归一化参数。注意，观测值 u_i 是历史研究 $j[i]$ 中国家 $k[i]$ 的国家启动时间。国家启动时间取决于多个因素的影响：国家索引（例如美国）、研究的治疗领域，以及可能的其他协变量。我们采用线性混合模型方法 [13,14,15] 对国家启动时间的对数进行建模，

$$\log(u_i) = \mu_{csu} + \gamma_{csu,k[i]} + \alpha_{csu} X_i + \epsilon_i. \tag{1}$$

μ_{csu} 是模型的总体截距。$\gamma_{csu,\ k[i]} \sim N(0, \sigma^2_{csu,\gamma})$ 表示从不同国家收集的国家启动时间的国家效应。$\gamma_{csu,\ k[i]}$ 被视为随机效应，因为一项研究通常会在数十个国家招募受试者。 X_i 表示第 i个研究-国家组合所属的治疗领域；由于在我们的情境中仅考虑四个治疗领域（即免疫学、肿瘤学、神经学和全科医学），因此 X_i 被视为固定因子。由于 X_i 有 4 个水平， X_i 的模型矩阵有 3 列（因为包含截距），而 α_{csu} 是一个包含 3 个元素的固定效应参数向量。最后，$\epsilon_i \sim N(0, \sigma^2_\epsilon)$ 表示独立同分布的正态随机误差。根据历史数据集以及投资组合规划时可获得的数据，还可以在模型中加入其他协变量。

如果认为不同历史研究中的国家启动时间差异很大，则可以纳入研究随机效应 $\beta_{csu,\ j[i]} \sim N(0, \sigma^2_{csu,\beta})$（与 $\gamma_{csu,\ k[i]}$ 无关），因为内部和外部数据库中通常有数十项历史研究可供建模使用。因此，国家启动时间模型可以表示为：

$$\log(u_i) = \mu_{csu} + \beta_{csu,j[i]} + \gamma_{csu,k[i]} + \alpha_{csu} X_i + \epsilon_i. \tag{2}$$

此外，如果图形绘制的探索性分析显示存在国家依赖的治疗领域效应，则可以纳入国家随机斜率，模型如下：

$$\log(u_i) = \mu_{csu} + \beta_{csu,j[i]} + \gamma_{csu,k[i]} + (\alpha_{csu} + \delta_{k[i]}) X_i + \epsilon_i, \tag{3}$$

其中， $\delta_{k[i]} \sim N(0, \sigma^2_\delta)$ 表示 X_i斜率中的随机效应。

虽然对数国家启动时间的线性混合效应模型具有计算时间短的优点，但这并不妨碍我们考虑广义线性混合效应模型 [14,16,17]，以直接对结果变量建模。例如，我们可以基于采用对数链接函数的广义线性混合效应模型，使用伽马分布对国家启动时间建模：

$$u_i \sim \text{Gamma}(\mu_i, \phi) \text{ with the Gamma distribution mean } \mu_i = \exp(\mu_{csu} + \beta_{csu,j[i]} + \gamma_{csu,k[i]} + \alpha_{csu} X_i), \tag{4}$$

其中， ϕ 是伽马分布的离散参数。

进行混合模型估计时，可使用 R 包 lme4 [18] 来估计线性混合效应模型和广义线性混合效应模型。在我们的案例中，研究和国家效应被建模为交叉随机效应，而 lme4 包提供了简洁的语法和高效的实现方式，用于拟合具有交叉随机效应的模型。

3.3 中心启动模型

在我们的问题中，我们观测到历史研究中每个国家所有已启动研究中心的启动日期。因此，可以使用贝叶斯分层非齐次泊松过程方法，对特定国家随时间变化的场地启动模式进行建模。例如，在 Wang 等 [12], 的研究中，该方法已用于根据截至某一时间点观测到的试验中数据，对特定研究中心内的受试者入组过程进行建模。在他们的应用中，用于模型估计的输入数据基于单项研究中的观测数据。我们将该方法扩展到多个历史研究，以纳入这些研究中观测到的招募相关数据。对于上一节定义的第 i个研究-国家组合， $j[i]$和$k[i]$ 分别是该 i个组合对应的研究索引和国家索引。 u_i，即该 i个组合的国家启动时间，是该 i个组合中心启动的起始时间。设 t'_i 为该 i个组合的场地启动结束日期，其中 t'_i 可选为在研究-国家层面或研究层面最后一个启动的研究中心的启动日期。采用研究-国家层面的中心启动结束日期，是基于场地启动期因国家而异这一假设；采用研究层面的中心启动结束日期，则基于这样一个假设：在已启动研究中心达到目标场地数量之前，所有国家的中心启动都不会停止。为适应研究的起始时间

点和所需的时间尺度，我们对 t_i' 进行标准化，得到 $u_i' = (t_i' - t_{0,j[i]})/m$，与处理 u_i 时相同。设 $N_{in,i}(u)$ 为第 i 个研究-国家组合在时间区间 $(u_i, u]$ 内启动的研究中心数量。请注意，首个启动的研究中心不计入 $N_{in,i}(u)$。随后，第 i 个研究-国家组合在 $(u_1, u_2]$ 内启动的研究中心数量可用泊松过程建模如下：

$$N_{in,i}\left(u_2\right) - N_{in,i}\left(u_1\right) \sim Poisson\left(\int_{u_1}^{u_2} \lambda_{in,i}(v)dv\right), u_2 > u_1 >= u_i. \tag{5}$$

如果认为研究中心启动率初始值较高，随后呈下降趋势，则可以使用以下时间衰减率函数：

$$\lambda_{in,i}^{(1)}(u) = \Lambda_{k[i]} \exp(-(\eta+\epsilon)(u-u_i)), \Lambda_{k[i]} \sim \Gamma(\alpha,\beta), \eta > 0, \epsilon = 10^{-5}(\text{offset}), u >= u_i. \tag{6}$$

基准研究中心激活率 $\Lambda_{k[i]}$ 因国家而异，并服从伽马分布。各国研究中心激活率的后验预测结果可为临床运营团队在不同国家进行研究中心规划与优化提供有意义的信息。如果探索性分析（例如按国家绘制研究中心启动日期的时间序列图）显示，研究中心激活率通常先升至峰值，然后逐渐下降，那么以下二次率函数将是不错的选择，

$$\lambda_{in,i}^{(2)}(u) = \Lambda_{k[i]} \exp\left(-\frac{(u-u_i-e)^2}{2(\eta+\epsilon)}\right), \Lambda_{k[i]} \sim \Gamma(\alpha,\beta), e > 0, \eta >= 0, \epsilon = 10^{-5}(\text{offset}), u >= u_i \tag{7}$$

二次率函数 $\lambda_{in,\ i}^{(2)}(u)$ 比时间衰减率函数 $\lambda_{in,\ i}^{(1)}(u)$ 更加灵活，因为 $\lambda_{in,\ i}^{(2)}(u)$ 当参数 e 接近 0 时，可以将衰减率函数作为一种特殊情况纳入其中。

对于第 i 个研究-国家组合，在时间区间 $(u_i, u_i']$ 内有关中心启动的观测数据包括已启动的研究中心总数，记为 $n_{in,\ i}\left(u_i'\right)$ 以及第 i 个组合对应的中心启动时间，即 $\overrightarrow{u_{in,\ i}} = [u_{in,i,m} | m \in \{1, ..., n_{in,i}\left(u_i'\right)\}]$。可以证明，对于所有 N 个采用时间衰减率函数的研究-国家组合，其中心启动数据的似然函数具有以下形式，

$$\begin{aligned} &L\left(\alpha,\beta,\eta;\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right) = \prod_{i=1}^{N} L\left(\alpha,\beta,\eta;\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right)\right) \propto \\ &\prod_{i=1}^{N}\left(\frac{\Gamma\left(n_{in,i}\left(u_i'\right)+\alpha\right)}{\Gamma(\alpha)\beta^{\alpha}}\left(\frac{\beta(\eta+\epsilon)}{\beta\left(1-\exp\left(-(\eta+\epsilon)\left(u_i'-u_i\right)\right)\right)+(\eta+\epsilon)}\right)^{\left(n_{in,i}\left(u_i'\right)+\alpha\right)} \prod_{m}\exp\left(-(\eta+\epsilon)\left(u_{in,i,m}-u_i\right)\right)\right) \end{aligned} \tag{8}$$

令 $\pi(\alpha)$、$\pi(\beta)$ 和 $\pi(\eta)$ 分别为参数 α、β 和 η 的先验分布。则后验分布为：

$$\pi\left(\alpha,\beta,\eta|\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right) \propto L\left(\alpha,\beta,\eta;\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right)\pi(\alpha)\pi(\beta)\pi(\eta) \tag{9}$$

以 α、β、η 和观测数据为条件，我们可以从以下伽马分布中生成国家 k 的 $\Lambda_k (k \in \{1, ..., C\})$：

$$f\left(\Lambda_k|\alpha,\beta,\eta;\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right) \sim Gamma\left(\sum_{i:k[i]=k} n_{in,i}\left(u_i'\right)+\alpha, \frac{\beta(\eta+\epsilon)}{\beta\left(1-\exp\left(-(\eta+\epsilon)\sum_{i:k[i]=k}\left(u_i'-u_i\right)\right)\right)+(\eta+\epsilon)}\right) \tag{10}$$

似然 基于二次率函数的函数及相应的后验分布为 ，

$$\begin{aligned} &L\left(\alpha,\beta,\eta,e;\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right) = \\ &\prod_{i=1}^{N}\left(\frac{\Gamma\left(n_{in,i}\left(u_i'\right)+\alpha\right)}{\Gamma(\alpha)\beta^{\alpha}}\left(\frac{\beta}{\beta\left(\sqrt{2\pi(\eta+\epsilon)}\left[\Phi(\frac{u_i'-u_i-e}{\sqrt{\eta+\epsilon}})-\Phi(\frac{-e}{\sqrt{\eta+\epsilon}})\right]\right)+1}\right)^{\left(n_{in,i}\left(u_i'\right)+\alpha\right)} \prod_{m}\exp\left(-\frac{\left(u_{in,i,m}-u_i-e\right)^2}{2(\eta+\epsilon)}\right)\right), \end{aligned} \tag{11}$$

$$\pi\left(\alpha,\beta,\eta,e|\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right) \propto L\left(\alpha,\beta,\eta,e;\overrightarrow{u_{in,i}}, n_{in,i}\left(u_i'\right), i \in N\right)\pi(\alpha)\pi(\beta)\pi(\eta)\pi(e), \tag{12}$$

其中， $\Phi(\cdot)$ 是标准正态分布的累积分布函数。以 α、 β、 η、 e 和观测数据为条件，我们可以从以下伽马分布中生成国家 k的 Λ_k，

$$f\left(\Lambda_k|\alpha,\beta,\eta,e,\overrightarrow{u_{in,i}},n_{in,i}\left(u_i'\right),i\in N\right)\sim Gamma\left(\sum_{i:k[i]=k}n_{in,i}\left(u_i'\right)+\alpha,\frac{\beta}{\beta\left(\sqrt{2\pi(\eta+\epsilon)}\sum_{i:k[i]=k}\left[\Phi\left(\frac{u_i'-u_i-e}{\sqrt{\eta+\epsilon}}\right)-\Phi\left(\frac{-e}{\sqrt{\eta+\epsilon}}\right)\right]\right)+1}\right). \tag{13}$$

在对中心启动率函数进行建模时，我们注意到可以考虑研究效应 $\Omega_{j[i]}$。与国家效应不同，研究效应对于模拟新试验并无直接意义，因为新的模拟试验总是不同于历史研究的另一项试验，而历史研究和未来研究使用的是一组相同的国家。鉴于加入额外研究效应会大幅增加计算复杂度，且其边际收益有限，我们选择不在中心启动建模中纳入该效应。

为计算模型参数的后验分布，我们使用 R 软件包 MCMCpack 中的 R 函数 MCMCmetrop1R[19]，该函数可从用户推导出的后验分布函数中生成样本（如文献 9 和 12 所示）。

3.4 受试者入组模型

在受试者入组过程的建模方面，我们的观测数据仅包括历史研究中受试者入组信息的中心层面汇总。对于历史研究中的每个研究中心，我们只能观察到入组受试者总数，而无法获知每位受试者实际入组的日期。这一数据局限使我们无法对各研究中心内受试者入组随机过程的时间动态进行建模，并将我们的方法限制为对总体中心层面入组率进行建模。

用于对入组进行建模的历史研究数据集包含来自 S项不同研究的多个 N 研究-国家组合；对于每个研究-国家组合 i，启动的研究中心数量为$n_{in,\ i}$。设$N_{en,\ i,\ m}$ 为第 i个研究-国家组合中研究中心 m 入组的受试者总数。则$N_{en,\ i,\ m}$ 可以用以下广义混合效应泊松回归模型[14,15,20],进行建模

$$N_{en,i,m}\sim\text{independently as }Poisson(\mu_{im})\text{ where }\log(\mu_{im})=\mu_{en}+\log(d_{im})+\gamma_{en,k[i]}. \tag{14}$$

参数 μ_{im} 表示第 i个研究-国家组合中研究中心 m 的平均入组受试者人数。 μ_{en} 是模型的总体截距。项 $\log(d_{im})$ 是偏置项，用于考虑第 i个研究-国家组合中研究中心 m 的入组时长，其中 d_{im} 为相应的总入组时长。此处，研究中心的入组时长定义为从中心启动日期到最后一名受试者入组日期之间的时间，并通过标准化因子 m进行标准化。随机参数 $\gamma_{en,\ k[i]}\sim N(0,\sigma^2_{en,\gamma})$ 表示国家效应， $k[i]$ 是第 i个研究-国家组合的国家索引。如果探索性图形分析表明，某个国家各研究中心的平均入组率在不同历史研究之间变化很大，则可以将研究随机效应$\beta_{en,\ j[i]}$ 添加到模型中。

值得注意的是，当参数 μ_{en} 被纳入 $\gamma_{en,\ k[i]}\sim N(\mu_{en},\sigma^2_{en,\gamma})$ 时，我们本质上就假设国家效应 $\lambda_{en,\ k[i]}=\exp(\gamma_{en,k[i]})$ 服从对数正态分布，其中 $N_{en,\ i,\ m}\sim Poisson(d_{im}\lambda_{en,k[i]})$。这揭示了我们提出的模型与常用的泊松-伽马模型之间的联系。在泊松-伽马模型中，我们会假设参数 λ_{en}服从伽马分布。

计算时，使用 lme4 包中的 glmer 函数[18]，通过纳入偏置项来估计混合效应泊松回归模型。

3.5 蒙特卡罗模拟：预测未来的招募过程

在继续预测当前规划中的研究未来招募时间线之前，需要估计三个入组相关模型（即国家启动时间模型、中心启动模型和受试者入组模型）的参数。我们采用基于蒙特卡罗模拟的方法，随机生成一定数量的模拟试验：对每个试验，先模拟三个招募相关模型的参数，再依次模拟各模拟试验中的国家启动时间、各模拟国家内研究中心的中心启动时间，以及各模拟国家中各模拟研究中心的受试者入组时间，直至预先设定的时间上限。我们选择模拟 1000 个试验。时间上限可设为某个组织内所有已完成历史试验的最长入组时长

，且可因治疗领域而异。计算复杂度很大程度上取决于该时间上限。因此，我们建议根据当前研究所属治疗领域的专家用户的经验来确定该上限，尤其是在需要探索一批入组情景时。在我们的案例中，我们将时间上限设为 5 年。对全部 1000 个模拟试验完成所有招募相关数据的模拟后，即可汇总任意感兴趣指标的中位数、较低和较高百分位数以及其他统计量。例如，为了提供研究入组完成时间预测，我们可以报告末例受试者首次给药日期的中位数及其较低和较高百分位数，作为预测区间。以下段落将详细说明如何从国家启动时间模型、中心启动模型和受试者入组模型中进行模拟。

要模拟新试验中的国家启动时间，我们需要模拟固定效应参数 μ_{csu} 和 α_{csu}、随机效应 $\gamma_{csu,\ k}$ 和 $\beta_{csu,\ j}$ 以及随机误差 ϵ_i的取值。对于固定参数 μ_{csu} 和 α_{csu}，我们从其限制性最大似然（REML）估计量的渐近正态分布中抽取数值，其中正态分布的均值等于 REML 估计值，标准差等于 REML 估计值的标准误。对于国家 $k \in \{1, ..., C\}$的随机国家效应 $\gamma_{csu,\ k}$，我们从正态分布中进行模拟，其均值和方差分别等于随机效应 $\gamma_{csu,\ k}$的条件众数和条件方差，因为新研究中使用的国家与历史研究中的国家相同。对于随机研究效应 $\beta_{csu,\ j}$，我们从其无条件分布$N(0, \hat{\sigma}^2_{csu,\beta})$ 中进行模拟，其中 $\hat{\sigma}^2_{csu,\ \beta}$ 是 $\sigma^2_{csu,\ \beta}$的 REML 估计值，因为正在规划的研究是一项不同于历史研究的新研究。对于随机误差 ϵ_i，我们从正态分布 $N(0, \hat{\sigma}^2_{\epsilon})$ 中进行模拟，其中 $\hat{\sigma}^2_{\epsilon}$ 是σ^2_{ϵ} 的 REML 估计值。模拟出所有参数后，可以通过例如方程 2 将它们组合起来，从而得到模拟试验 j中国家 k 的国家启动时间。

为进行中心启动模拟，我们首先需要从 9 或 12 的后验分布对应的马尔可夫链蒙特卡洛样本中抽取模型参数（即 α、 β、 η 和 e）的取值。根据先前抽取的模型参数值和观测到的历史数据，从 10 或 13 的伽马分布中抽取国家特异性的中心激活率基准值 Λ_k。随后，按照 5 所定义的非齐次泊松过程，抽取待启动的中心总数及其启动时间。

进行受试者入组模拟时，我们采用与国家启动时间模拟类似的方法，为模拟试验 j 中的国家 k 模拟固定效应参数 μ_{en} 和随机效应 $\gamma_{en、\ k}$ 和 $\beta_{en、\ j}$ 的取值。入组时长偏移量根据预先设定的模拟时间上限（5 年）确定。随后，按照齐次泊松过程抽取待入组受试者总数及其入组时间。

4 性能评估

4.1 一般考量

为验证所提出建模框架的性能，我们从本机构中选取一组近期已完成的研究，运行模型以生成入组时长预测，并将预测结果与真实值进行比较。验证研究的入组时长定义为方案批准日期与末例受试者首次给药日期之间的时间间隔。要对验证研究进行入组预测，提供给模型的输入包括该验证研究的受试者总数、中心总数、治疗领域、疾病适应症和患者人群。此外，还使用与验证研究具有相同疾病适应症和患者人群的历史研究的入组相关数据来估计模型组件。此处，历史研究是指在验证研究方案批准日期之前已完成入组的研究。我们将所提出建模框架的性能与先前内部开发的入组预测系统进行比较；该系统基于传统统计方法，拟合简单的概率分布，将在下一小节中介绍。此外，我们通过比较不同名义水平预测区间的名义水平与覆盖率，展示我们的建模与模拟框架能够正确校准数据变异性。最后，我们展示如何生成随时间变化的预测入组曲线，并叠加置信带；用户认为这些曲线对试验运营规划非常有参考价值。

4.2 一种先前的入组预测系统

在先前内部开发的一种入组预测系统中，整个入组流程也被划分为三个定义各不相同的阶段。第一阶段称为中心启动期-1，指从最终方案批准

表 1 模型性能验证候选研究集的筛选标准

标准	值
入组完成日期	过去 3.5 年内
干预性研究与观察性研究	仅干预性研究
研究阶段	2 和 3
治疗领域	肿瘤学、神经科学、免疫学和全科医学
中心总数	10-500 个研究中心
受试者总数	50-1600 名受试者

表 2 不同治疗领域中入选研究的数量

	免疫学	全科医学	神经科学	肿瘤学
研究数量	14	4	3	4

日期到各研究中心开放日期之间的时段。第二阶段为研究中心启动期-2，定义为研究中心开放日期至该中心首例受试者入组日期之间的时长。最后一个阶段是入组期，指从该中心首例受试者入组日期到整个研究最后一名受试者入组日期之间的时段。与我们提出的建模框架相比，先前预测系统所依据的模型较为基础。对于中心启动期-1，未采用正式的统计建模；模拟仅通过自助法，从相同治疗领域和国家的历史数据中有放回地抽取已观察到的时长。对于研究中心启动期-2，建模使用对数正态分布拟合各国家的历史数据，然后从中随机抽样，供各国的不同研究中心使用。对于受试者入组的建模，则采用国家层级建模方法。针对每个国家，将伽马分布分别拟合至该国所有研究中心历史观察到的入组率；随后在模拟阶段，从伽马分布中随机抽取各研究中心的入组率，并将其用作泊松分布的参数，以模拟每天的受试者入组情况。

4.3 评估研究的选择

为确保研究选择过程的客观性，我们成立了一个独立于建模团队（作者所属团队）的指导委员会，负责主导选择一组已完成入组的研究用于模型评估。首先，我们根据表 1 所列的搜索条件，从内部数据库中筛选出候选研究集。随后，我们绘制每项候选研究的实际入组曲线，以检查是否存在暂停或入组速度剧烈变化等异常形态。这些异常形态通常是由罕见且不可预见的事件造成的。每当发生此类事件时，我们的内部用户都会逐个评估其影响，因此我们不应期望任何模型仍能提供准确的预测结果。根据对入组曲线的人工检查，我们提出一份入组曲线形态异常的研究清单，并建议将这些研究从模型性能评估中剔除；指导委员会则最终决定某项研究是否应列入评估研究清单。最终，我们从四个不同的治疗领域中选出 25项研究，以测试所提出的建模方法，并与先前内部开发的预测系统进行比较。表 2 展示了这 25项研究在四个治疗领域中的分布。

4.4 预测准确度

为评估模型性能，我们将所提出的建模框架和先前的入组预测系统应用于独立指导委员会选出的25项研究。就性能评估指标而言，我们报告真实入组时长落入半径固定为 +/-1、2和3个月的预测区间内的覆盖率（因此，预测区间的宽度分别为2、4和6个月），以及真实入组时长落入

表3 我们的新建模框架与先前内部开发的入组预测系统对25项研究的入组时长预测

模型	覆盖率 2个月预测区间	覆盖率 4个月预测区间	覆盖率 6个月预测区间	覆盖率 95%预测区间	宽度 95%预测区间	绝对中位数预测误差	绝对均值预测误差
New	7（28%）	8（32%）	13（52%）	23（92%）	20个月	2.8个月	4.8个月
先前	5 (20%)	6 (24%)	9 (36%)	5 (20%)	2 个月	5.6个月	5.8个月

表4 我们新建模框架所得预测区间的名义水平与实际覆盖率

预测区间的名义水平	40%	50%	70%	95%
正确预测的研究数量	13	15	17	23
覆盖率	52%	60%	68%	92%
覆盖率的标准误	10%	10%	9%	5%
覆盖率的95%置信区间	32%-72%	40%-80%	50%-86%	82%-100%
预测区间宽度	5.3个月	7个月	10个月	20个月

名义置信水平为95%的非固定宽度预测区间内的覆盖率。此外，我们还考察预测入组时长中位数与真实值相比的绝对预测误差均值和中位数。

从表3可以看出，与先前的入组预测系统相比，我们提出的方法在各种固定宽度的所有预测区间上都取得了更高的覆盖率。在中位数绝对预测误差方面，我们提出的方法将误差从5.6个月降低至2.8个月，减少了50%；而对于通常受较大误差影响的平均绝对预测误差，我们的方法仍将误差从5.8个月降低至4.8个月，减少了17%。这表明，与我们基于混合效应模型和非齐次泊松过程模型的更复杂建模方法相比，传统的基于分布的方法不足以捕捉招募相关过程背后的复杂模式。表5和表6分别详细列出了我们新的建模方法和先前预测系统的输出结果。

4.5 预测区间的覆盖率

对于名义水平为 95% 的预测区间，我们的方法实现了 92% 的实际覆盖率（见表3）；而先前的预测系统校准不佳，覆盖率仅为 20%，且区间宽度极窄。这表明，先前的入组预测系统在建模或模拟过程中都未能正确考虑历史数据的变异性。由于先前的预测系统未提供其他名义水平的预测区间，我们进一步考察了所提出的新建模方法在不同名义水平下的名义覆盖率与实际覆盖率之间的差异。表4显示，在不同名义水平下，名义水平与实际覆盖率非常一致。这是因为我们的方法学通过适当的随机效应，正确考虑了不同层级（如研究层级和国家层级）的数据变异性。在模拟阶段，相应地从正确的条件或无条件渐近抽样分布中模拟随机效应（见第3.5小节）。尽管 95% 是统计假设检验中常用的显著性水平，但在我们的应用场景中，95% 预测区间的宽度通常过大，难以实际使用。这往往是因为入组相关过程本身存在大量变异性，而且可用于模型估计的历史研究数量有限。在我们的应用中，40% 或 50% 的预测区间通常能较好地平衡覆盖率与区间宽度。

4.6 预测入组曲线

除了预测入组持续时间之外，用户还发现，获取从方案批准到入组完成整个时间段内的预测入组曲线，对他们的规划工作非常有帮助。这可以通过以下方式实现：

表5 我们新的建模框架对25项研究入组时长的预测结果（单位：月）

研究 NO.	实际值 持续时间	预测中位数 持续时间（95%预测区间）	残差	以内 +/- 1 月	在范围内 +/- 2 个月	内部 +/- 3 个月
1	21.23	23.94（14.66，44.90）	2.70	No	No	Yes
2	20.93	20.26（14.02，34.03）	-0.67	Yes	Yes	Yes
3	21.87	15.17（10.38，22.68）	-6.69	No	No	No
4	15.63	16.16（9.88，30.08）	0.53	Yes	Yes	Yes
5	22.47	29.22 (11.45 , Inf)	6.75	No	No	No
6	26.37	20.46（13.71，31.27）	-5.90	No	No	No
7	23.20	23.63（17.58，34.35）	0.43	Yes	Yes	Yes
8	26.83	24.31 (16.07 , 40.46)	-2.53	No	No	Yes
9	30.27	20.91（10.95，51.80）	-9.35	No	No	No
10	37.33	25.31（15.37，55.11）	-12.02	No	No	No
11	28.77	26.02 (17.11 , 40.29)	-2.74	No	No	Yes
12	19.93	19.18（12.29，34.41）	-0.75	Yes	Yes	Yes
13	17.53	20.18（13.54，35.40）	2.65	No	No	Yes
14	11.27	10.13（6.38，17.34）	-1.14	No	Yes	Yes
15	10.03	10.07（6.89，17.29）	0.03	Yes	Yes	Yes
16	21.13	22.11（15.72，36.70）	0.98	Yes	Yes	Yes
17	19.93	12.77 (7.96 , 20.31)	-7.16	No	No	No
18	16.77	19.55（9.96，60.24）	2.79	No	No	Yes
19	18.37	14.86（9.24，25.40）	-3.50	No	No	No
20	30.80	19.14 (12.48 , 34.99)	-11.66	No	No	No
21	24.90	18.61（9.39，46.12）	-6.29	No	No	No
22	26.37	31.81 (21.82 , 51.40)	5.44	No	No	No
23	20.63	11.58（6.90，17.76）	-9.05	No	No	No
24	16.10	35.03（16.21，Inf）	18.93	No	No	No
25	10.83	11.58（8.11，17.70）	0.75	Yes	Yes	Yes

首先选取一组间隔均匀的时间点（例如，以月为单位），然后汇总网格中每个时间点的入组总人数中位数、较低百分位数和较高百分位数。在图2和图3中，我们展示了对预测效果良好的研究得出的预测入组曲线及其95%和50%置信带。请注意，根据用户的偏好，置信带上限不超过该研究的目标受试者入组人数。

5 讨论

针对规划阶段的入组预测问题，我们开发了一种新颖的统计建模框架。该框架基于广义线性混合效应模型（GLMM），并通过贝叶斯框架采用非齐次泊松过程，以系统地分阶段建模国家启动、中心启动和受试者入组。基于来自四个治疗领域的25项预先选定研究，我们的新建模框架与拟合简单概率分布的传统统计方法相比，显著提高了预测准确度。此外，我们还表明，该建模与模拟框架能够恰当地校准数据变异性，并为具有不同名义水平的预测区间提供正确的覆盖率。通过蒙特卡罗模拟，我们演示了如何生成带有置信带的中位数预测入组曲线。要对其他与招募相关的指标进行预测也并不困难，例如每个国家的研究中心中位数以及国家启动日期的中位数。

在国家启动时间建模方面，我们采用广义线性混合效应模型，以考虑历史研究中的国家层面和研究层面变异，以及固定协变量的影响。机器学习模型

表6 先前内部开发的入组预测系统对25项研究入组时长的预测结果（单位：月）

研究 NO.	实际值 持续时间	预测中位数 持续时间（95%预测区间）	残差	内部 +/-1 月	内部 +/-2 个月	内部 +/-3 个月
1	21.23	24.07（22.20，26.10）	2.83	No	No	Yes
2	20.93	21.37 (19.90 , 22.83)	0.43	Yes	Yes	Yes
3	21.87	13.53（12.87，14.17）	-8.33	No	No	No
4	15.63	14.23（13.47，14.93）	-1.40	No	Yes	Yes
5	22.47	19.00（18.03，20.13）	-3.47	No	No	No
6	26.37	17.67（16.87，18.37）	-8.70	No	No	No
7	23.20	23.70（22.50，24.70）	0.50	Yes	Yes	Yes
8	26.83	14.33（13.47，15.20）	-12.50	No	No	No
9	30.27	26.50（24.37，28.63）	-3.77	No	No	No
10	37.33	25.80（24.47，27.60）	-11.53	No	No	No
11	28.77	16.30（15.73，16.83）	-12.47	No	No	No
12	19.93	20.40 (19.00 , 22.17)	0.47	Yes	Yes	Yes
13	17.53	13.50（12.67，14.23）	-4.03	No	No	No
14	11.27	9.07 (8.47 , 9.90)	-2.20	No	No	Yes
15	10.03	9.60 (8.83 , 10.40)	-0.43	Yes	Yes	Yes
16	21.13	15.33（14.47，16.10）	-5.80	No	No	No
17	19.93	13.43（12.63，14.13）	-6.50	No	No	No
18	16.77	14.40（13.37，15.43）	-2.37	No	No	Yes
19	18.37	12.77 (11.73 , 13.63)	-5.60	No	No	No
20	30.80	19.90（18.83，21.23）	-10.90	No	No	No
21	24.90	32.07 (29.07 ， 36.37)	7.17	No	No	No
22	26.37	34.73（30.27，39.97）	8.37	No	No	No
23	20.63	11.40（10.63，12.20）	-9.23	No	No	No
24	16.10	31.07 (27.37 , 35.00)	14.97	No	No	No
25	10.83	11.73（10.80，12.70）	0.90	Yes	Yes	Yes

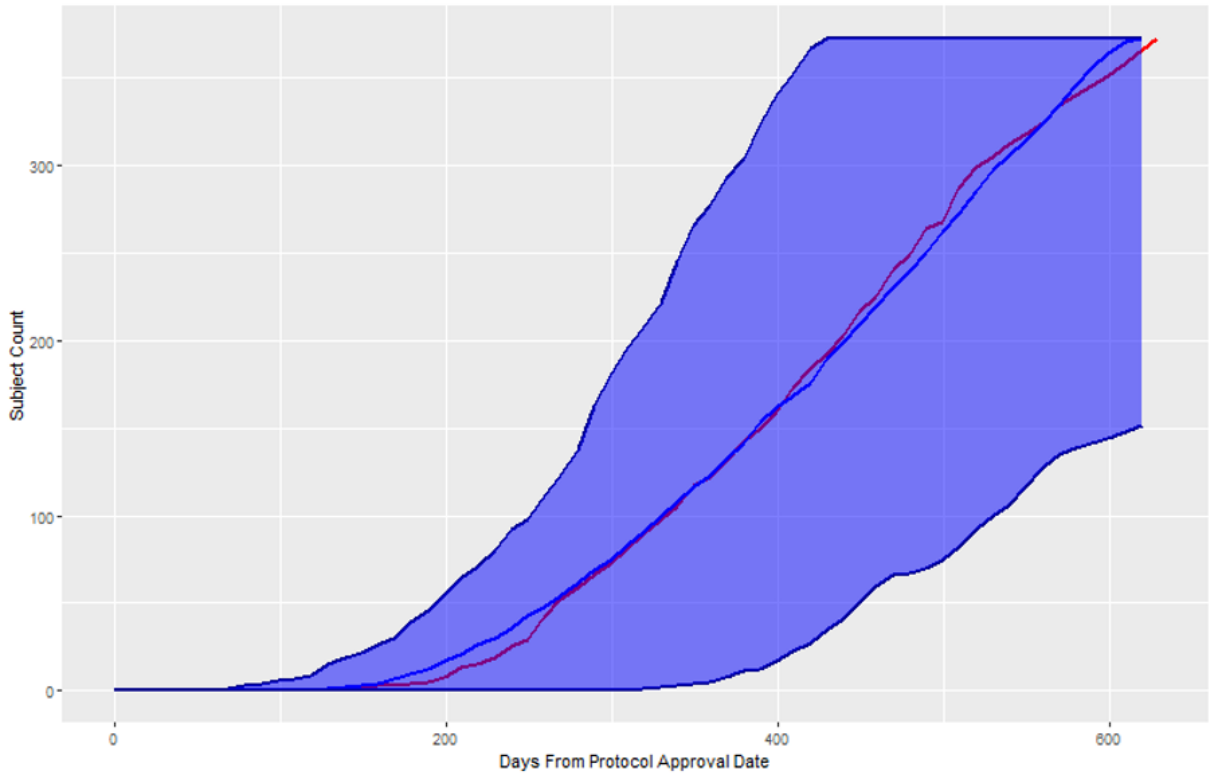
（例如，梯度提升）可用于在有更多协变量可供建模时进行预测。但要恰当地考虑机器学习模型预测结果的变异并非易事。另一个潜在的局限是，进行入组预测时需要更多协变量信息。对于处于投资组合规划阶段的研究，通常掌握的研究信息有限。

关于国家内研究中心启动过程的建模，如果能够获得特定国家的启动日期，我们可能需要考虑这样一个事实：从国家启动日期到首个研究中心启动日期的时长，可能远长于随后启动的研究中心之间的时间间隔。在这种情况下，我们建议将国家启动至首个研究中心启动之间的时间与其余已启动研究中心的启动过程分开建模。另一种方法是将国家启动至首个研究中心启动之间的时间纳入国家启动时间的定义中，并在国家启动时间的建模中隐式处理。

对于受试者入组的建模，我们的观测数据仅包含历史研究中中心层面的受试者入组汇总信息。如果观测到每位受试者的实际入组日期，则可以使用非齐次泊松过程，估计各国家研究中心内受试者入组过程的时变速率函数。同样，我们可能需要考虑到，由于研究中心准备工作的开销，首例受试者入组通常需要更长时间，因此单独对首例受试者入组所需时间进行建模可能更为合适。

使用依赖历史研究进行模型估计的入组预测系统时，最终用户（例如临床运营团队）务必对所用的历史研究进行数据核查。如果历史研究中有关入组速度等汇总统计量系统性地低于或高于用户的预期，则在解读预测入组持续时间时需要谨慎

图2带有95%置信带的预测入组曲线



。可以考虑采取某些临时措施，例如排除一些具有极端统计值的研究，或使用乘数调整入组速度。

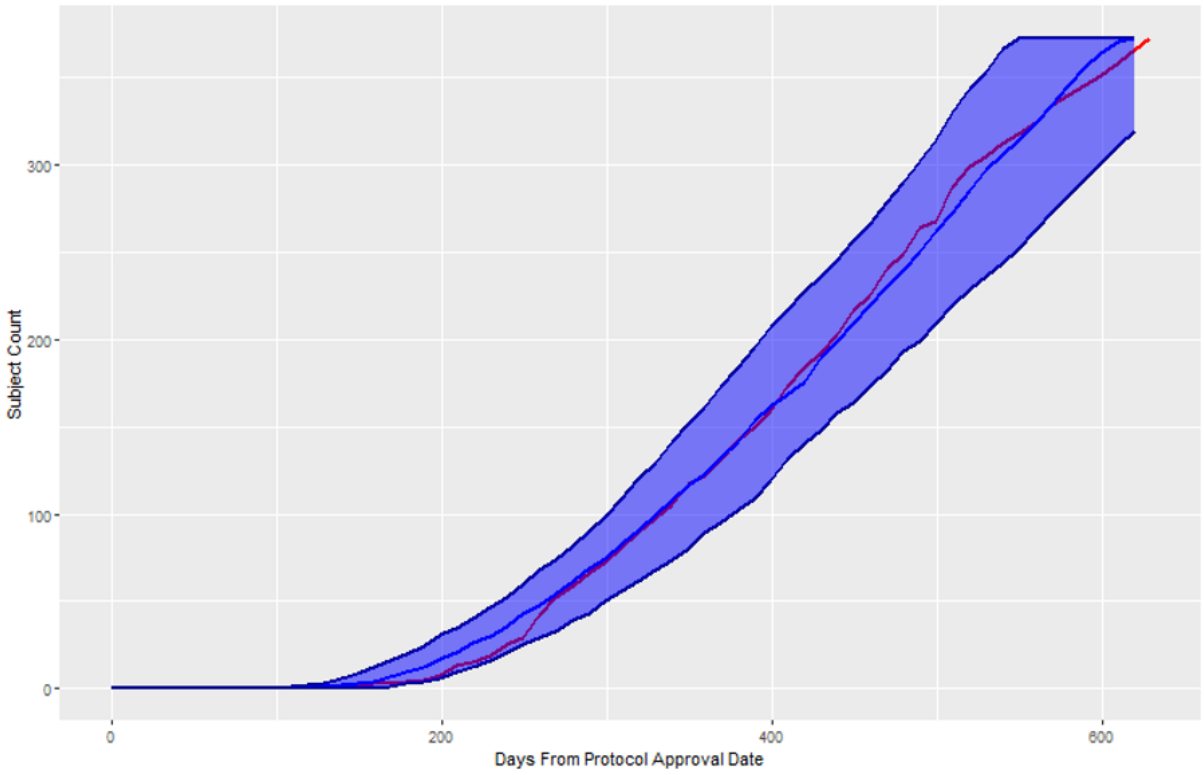
披露

本手稿由艾伯维赞助。艾伯维参与了研究设计、研究实施、数据解读、撰写和审阅，并批准了内容。所有作者均为艾伯维公司的员工，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。

图3 带有40%置信带的预测入组曲线



参考文献

1. 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。十周年特刊。doi：https://doi.org/10.1016/j.cct.2015.07.010 2. Anisimov VV。关于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。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。多中心试验中的建模、预测和招募自适应调整。*Statisticsin Medicine* 2007年；26(27)：4958–4975。doi：https://doi.org/10.1002/sim.2956 5. 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）。6. Anisimov VV。临床试验的统计建模（招募和随机化）。*Communications inStatistics-TheoryandMethods*2011年；40(19-20)：3684–3699。7. 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）。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. 临床试验入组的随机建模与预测。*Statistics inMedicine* 2010年；29(6)：649–658。11. Zhang X, Long Q. 临床试验入组与事件时间的联合监测和预测。《生物统计学杂志》2012年；54(6)：735–749。12. Wang L, Liu Y, Chen X, Pulkstenis. 临床试验中终点成熟时间的实时监测与预测。*Statistics in Medicine*2022年；41(18)：3596–3611。13. Davison AC. 统计模型。11. 剑桥大学出版社。2003年。14. McCulloch CE, Searle SR. 广义线性与混合模型。约翰·威立父子公司。2004年。15. McCullagh P, Nelder JA. 广义线性模型。Routledge。2019年。16. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. 纵向数据分析。CRC出版社。2008年。17. Goldstein H, Browne W, Rasbash J. 多层模型中的变异分解。《理解统计学：心理学、教育学和社会科学中的统计问题》2002年；1(4)：223–231。18. Bates D, Mächler M, Bolker B, Walker S. 使用 lme4 拟合线性混合效应模型。《统计软件杂志》2015年；67(1)：1–48。doi：10.18637/jss.v067.i01 19. Martin AD, Quinn KM, Park JH. MCMCpack：R 中的马尔可夫链蒙特卡罗。《统计软件杂志》2011年；42(9)：22。doi：10.18637/jss.v042.i09 20. Austin PC, Stryhn H, Leckie G, Merlo J. 率/计数数据泊松回归多层分析中的聚类与异质性测量。*Statistics in medicine* 2018年；37(4)：572–589。

