

T/CNPHARS

中 国 药 理 学 会 团 体 标 准

T/CNPHARS ****—****

群体药动学-药效学分析标准操作规范

Standard operation procedure for
population pharmacokinetic-pharmacodynamic analysis

（征求意见稿）

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中国药理学会 发 布

目 录

目录

前 言	III
引 言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
3.2 基础模型	1
3.3 最终模型	1
3.4 模型评价	1
4 符号	1
5 标准操作规范	2
5.1 概述	2
5.2 前期工作	3
5.2.1 分析计划	3
5.2.2 分析数据集	3
5.2.3 探索性数据分析	5
5.3 基础模型	6
5.3.1 定义	6
5.3.2 算法	6
5.3.3 结构模型的建立	7
5.3.4 随机效应模型的建立	8
5.4 最终模型	9
5.4.1 定义	9
5.4.2 协变量模型	10
5.4.3 随机效应模型	11
5.4.4 其他	11
5.5 模型评价	13
5.5.1 定义	13
5.5.2 一般原则	13
5.5.3 诊断图评价	13
5.5.4 数值统计评价	17
5.6 模型应用	17
5.6.1 估算个体 PK 参数和暴露量	17
5.6.2 考察协变量的影响	18
5.6.3 模拟	18
5.7 分析报告	18
5.7.1 摘要	19
5.7.2 研究背景	19
5.7.3 目的	19
5.7.4 数据	19
5.7.5 分析方法与假设	19
5.7.6 结果	19
5.7.7 讨论	19
5.7.8 结论	19
5.7.9 附录	20

5.7.10 其他.....	20
5.8 质量控制.....	20
5.8.1 定义.....	20
5.8.1 一般流程.....	20
附录 A 质量控制要点.....	23
A1 数据.....	23
A2 控制文件、输出文件和数据后处理.....	25
A3 分析报告.....	27
参考文献.....	29

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国药理学会定量药理学专业委员会提出。

本文件由中国药理学会归口。

引 言

群体药动学-药效学分析是应用数学建模的方法，将经典药动学-药效学理论与统计学原理和方法相结合，考察研究对象中药动学和药效学的定量关系、以及内在和外在的影响因素和随机效应。与经典的药动学-药效学分析相比，群体药动学-药效学分析可充分利用稀疏采样的数据进行分析，有利于在患者人群、尤其是特殊人群如老年人、儿童、孕产妇和危重症等患者中开展研究。

群体药动学旨在定量描述药物在体内吸收、分布、代谢和排泄的过程，并定量考察药动学的变异大小及其来源。而群体药效学旨在定量描述药物效应随药物暴露量和时间变化过程，并考察药物效应的变异程度及其来源。两者相结合，可定量描述药物、机体和疾病之间的关系，为药物研发和临床合理用药决策提供重要依据，是“模型引导的新药研发（model informed drug development）”和“模型引导的精准用药（model informed precision dosing）”中的重要分析技术。

群体药动学-药效学分析是一个学习和确认的迭代过程，涉及了多学科的理论 and 知识，分析过程复杂且耗时，亟待对分析流程进行规范化和标准化，以提高分析效率和分析质量。本规范的内容主要基于当前对群体分析的理解和认识，提供相关的科学指导，以帮助合理开展和应用群体分析。未来随着学科不断发展，需基于科学判断开展研究和分析，并更新本规范。

2

3

群体药动学-药效学分析标准操作规范

1 范围

本文件提供了群体药动学-药效学分析的内容、流程和相关要求。

本文件适用于指导新药研发和临床合理用药中开展群体药动学-药效学分析的工作。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1 群体药动学-药效学

群体药动学-药效学是指应用数学建模的方法，将经典药动学-药效学理论与统计学原理和方法相结合，考察和建立研究对象中药动学和药效学的定量关系，以及定量分析内在和外在的相关影响因素和随机效应。

3.2 基础模型

基础模型表征数据整体特征的模型，包括结构模型与统计学模型两部分。

3.3 最终模型

最终模型是经过筛选所有计划筛选协变量后保留所有相关协变量的最精简模型。

3.4 模型评价

模型评价为客观评估模型在特定领域的预测能力，或判断模型偏倚是否会对决策产生实质性的影响。

4 符号

下列符号适用于本文件。

BLQ: 低于定量下限数据

DV: 因变量

EDA: 探索性数据分析

ETA: 个体间变异

EPS: 残差变异

IPRED: 个体预测值。

NONMEM: 非线性混合效应模型计算机软件

OBS: 观测值

PRED: 群体预测值

PK: 药动学

PK-PD: 药动学药效学

PMAP: 群体建模分析计划

PMAR: 群体建模分析报告

PPK: 群体药动学

PPD: 群体药效学

PPK-PD: 群体药动学-药效学

OFV: 目标函数值

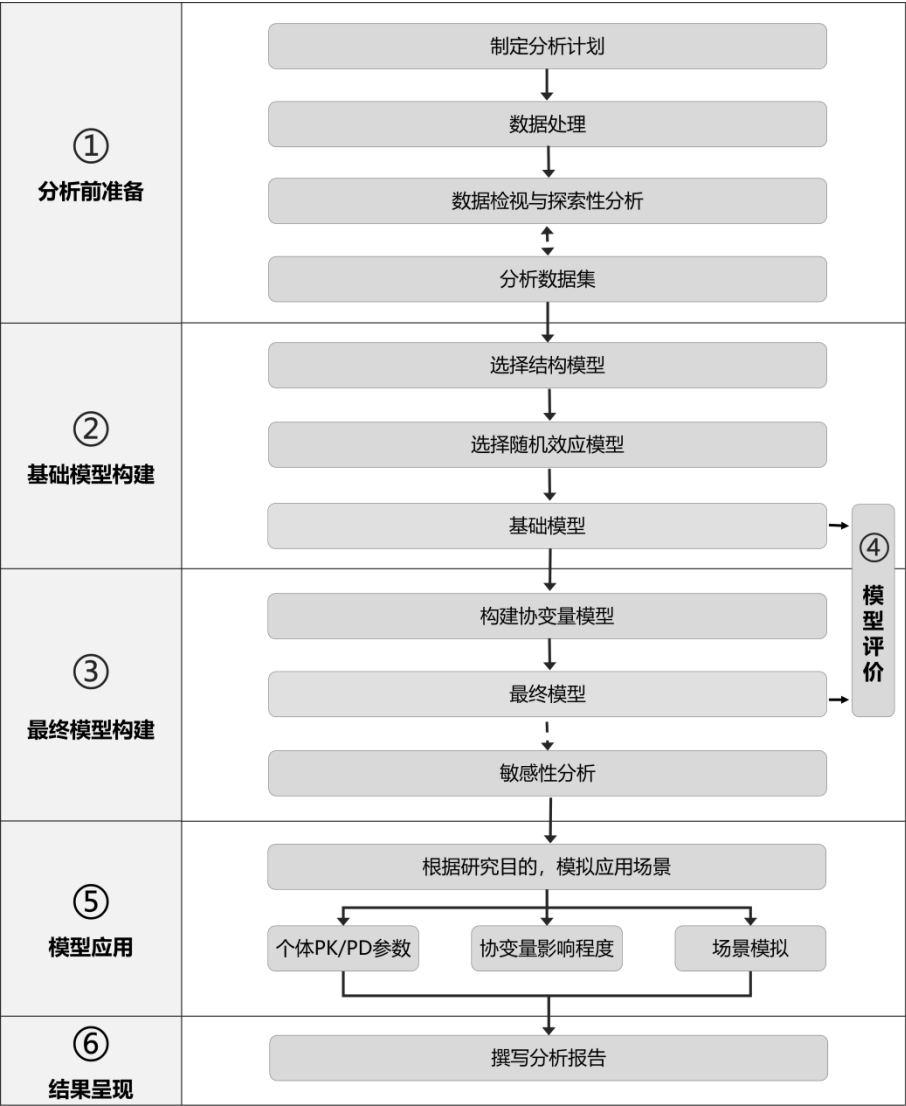
VPC: 可视化预测检验。

42

5 标准操作规范

5.1 概述

群体药动学-药效学分析主要涵盖以下六大步骤：分析前准备、基础模型、最终模型、模型评价、模型应用和分析报告，见下图。此外，质量控制应贯穿整个分析过程，确保分析全过程的准确、可靠和可追溯。



群体药动学-药效学分析流程图

虚线箭头代表在特定情况下需完成的步骤，双向箭头则表示步骤间可能会发生的循环往复过程。

鉴于群体药动学-药效学分析是一个持续的“学习与确认”（learn and confirm）的迭代过程，故可根据分析目标，分为“探索型”和“决策型”两大类。其中，“探索型”分析通常是不断积累数据和探索“药物—人体—疾病”相互作用规律的过程；而“决策型”分析一般指对药物研发或临床用药等决策产生重大影响的分析。

不同分析类型的一般分析流程和评价标准存在差异。“探索性”分析可根据其分析目标对流程进行合理调整。相较于“决策型”分析，“探索性”分析具有更高的灵活性。而“决策型”分析则应遵循既定的规范和流程系统地进行。应根据已有的分析数据和分析目标，明确分析类型和遵循的规范。除非特别说明，下文均指“决策型”分析的标准规范流程。

5.2 前期工作

开展 PPK-PD 分析之前，应事先制订相应的群体建模分析计划（Population Modelling Analysis Plan, PMAP），说明研究背景、分析目标、拟纳入分析的临床研究项目、分析数据集的创建、模型化分析策略、拟构建的模型特征和模型模拟应用等，并同时制定分析数据的说明文件，以便数据管理人员准备相应的分析数据集。

5.2.1 分析计划

通常，分析计划应包括以下内容：

- (1) 研究背景：提供药物研发的背景信息，包括药物的作用机制、拟开发的适应症、研发的阶段、相关的疾病和药理学/药动学信息、既往相关群体分析的总结等。
- (2) 分析目的：待解决的药物研发或临床实践中的问题。
- (3) 分析目标：包括且不限于描述药物在研究对象中的 PK 或 PK-PD 关系、分析协变量对 PK 或 PK-PD 的影响、考察药物暴露量和疗效终点的关系、探索暴露量和安全性的关系等。
- (4) 研究假设：定量描述 PK 特征，以及 PK-PD 之间的直接、间接或其他效应关系。
- (5) 数据来源：详细介绍拟纳入研究项目的名称、研究设计、研究对象及其特征、给药方案、PK-PD 采样方案、生物样本分析方法等相关信息。
- (6) 数据准备：阐述源数据集的格式（如 SDTM、ADaM 或其他 CDISC 格式）及其转换（如转换为.csv 格式）、数据集变量的定义、数据集修改更新的标识、数据集质量控制规范和要求、数据集应包括的内容（如剂量、浓度-时间数据、人口学数据、探索的协变量信息、PD 和临床效应相关的信息等，并提供详细的数据集内容说明）、数据的处理办法（如缺失数据、低于定量下限数据和离群数据的处理方法等）。
- (7) 分析软件：如 NONMEM、Phenoix NLME、Monolix、R、R package、SAS 等。
- (8) 分析方法：包括数据探索性分析、基础模型构建、协变量分析、最终模型构建，模型选择标准、模型评估方法等。此外，应对可能影响最终模型结论的假设进行详细说明，但不应限制分析过程中发生的假设内容补充、替换或删除等，但假设的改动须在最终研究报告中进行阐述和评价。
- (9) 模型应用：计划应用模型解决的问题，如评估协变量的影响、模拟不同给药方案、比较不同人群的暴露量等。

鉴于药物研发的不同阶段对药物 PK-PD 的认知理解程度不同，PMAP 的撰写策略应相应地进行调整。在药物研发的早期阶段，由于对药物 PK 和 PD 的认识相对有限，PMAP 可被视为一个动态的“活文件”。在此阶段，分析方法的描述可适当简化，并应随着研发过程中对药物 PK 和 PD 理解的逐步深入而进行相应的补充和优化。在此阶段，PMAP 应重点明确分析目标和旨在解决的关键研发或临床问题，无需过多叙述技术细节。同时，分析人员在分析过程中应当准确记录 PMAP 的任何修订。在药物研发的晚期阶段，尤其是针对确证性临床试验等前瞻性研究的分析，以及用于指导药物开发和与监管机构沟通的分析，PMAP 应在数据锁定之前确定，并清晰阐述分析过程中的关键假设和模型评价标准。

表 1 准备和提交群体模型分析计划的时间节点

分析目的	准备和提交的时间节点
<ul style="list-style-type: none"> 药物研发中的探索性和回顾性分析 学术研究 指导个体化精准用药 	开展分析前完成初步的分析计划，在分析过程中可根据分析目的、数据的可及性等进一步调整、更新计划。
<ul style="list-style-type: none"> 药物研发后期的前瞻性研究（如确证性临床研究） 	研究分析前定稿。
<ul style="list-style-type: none"> 用于药物开发和监管的重要决策（如采用分析结果替代临床研究、基于 PK 暴露量的外推等） 	开展研究前(包括 IND, EOP1/2,或沟通交流阶段)提交监管部门评审，以寻求共识。

5.2.2 分析数据集

在开展群体分析时，纳入的数据取决于建模的目的。通常情况下，临床早期试验收集密集采样的数据，可用于建立准确的基础结构模型。而在晚期临床试验中，通常收集了稀疏采样的数据。这些数据包

含了重要的协变量信息，适合用于建立协变量模型。PMAP 需明确说明为达到分析目的所需的具体分析数据。

5.2.2.1 数据整理

建议遵循临床数据交换标准协会(CDISC)制定的、适用于群体分析、且符合分析数据模型(ADaM)标准的基本数据结构实施指南 (ADaM popPK Implementation Guide v1.0 | CDISC, <https://www.cdisc.org/adam-poppk-implementation-guide-v1-0>)。如果未使用该指南，则应制定数据收集表格(Data Request Form, DRF)，明确所用的变量，便于数据整理和编辑。在制订 DRF 时，应全面考虑数据分析的需求，纳入后续分析所需的指示变量和时间变量等、使后续的数据整理工作更高效便捷。

5.2.2.2 完整数据与验证数据

建模过程中，建议采用完整数据 (full data-set)，以包括尽可能多的信息。这有利于建立稳健的模型。模型评估过程中，可用外部验证数据来考察模型的预测效能。外部验证数据指建模过程中未纳入的数据，或是其他研究产生的数据。

5.2.2.3 低于定量下限的数据

在定量检测下限以下的数据称为低于定量下限 (below Limit of Quantification, BLQ) 数据。由于不同的处理方式可显著影响参数估算值，故应谨慎合理地处理 BLQ 数据。例如，临床疗效的应答率直接从 PK 或 PD 数据衍生而来时，如应答率定义为 PK 或 PD 数值高于 (或低于) 某一临界值，则将任何 BLQ 数据排除在群体分析之外，均将错误估算临床疗效应答率。故在此情况下，应将所有 BLQ 数据纳入分析之中。但是，如果 BLQ 数据的占比 $\leq 10\%$ ，且无系统性趋势分布，则一般可直接忽略这些 BLQ 数据。

当 BLQ 数据的占比超一定比例时，如 $>10\%$ 时，应首先排查研究设计、生物检测、受试者依从性等方面的影响。如果排查后认为数据仍可用于分析，则应评估处理方法对模型参数估算的影响。BLQ 数据的占比与其对应的处理方式须在分析计划中事先约定。通常可首先考虑直接舍去 BLQ 数据 (M1 法)，然后考察 BLQ 数据的分类变量可视化预测检验图，如预测和观测的 BLQ 比例一致性较好，则可以舍去 BLQ 数据。若一致性不佳，可将每个受试者的首个 BLQ 数据设为定量下限的一半，舍去余下的 BLQ 数据 (称为 M6 方法)，再观察预测 BLQ 和观测 BLQ 比例的一致性。如一致性较好，则采用该方法。若一致性不佳，则在估算模型参数的同时，估算 BLQ 数据的似然 (称为 M3 方法)。同时，建议采用敏感性分析等方法，证明 BLQ 数据处理方法的合理性。

5.2.2.4 缺失数据

如果出现缺失数据，应尽可能查找数据。但有时缺失数据难以复得，须对缺失数据进行合理和科学的处理。

缺失数据可包括缺失浓度数据、缺失时间数据、缺失给药数据和缺失协变量数据。本部分仅讨论前三种缺失情况，缺失协变量数据将在后文中讨论。

(1) 缺失浓度数据应被直接舍弃。

(2) 缺失采样时间时，由于无法计算给药后时间，故应在试验中及时进行数据核查减少此类错误。在试验过程中，时间应采用 24 小时制记录、并按照日期和时间的顺序记录。如有缺失数据，应在病例报告表中详细记录。

(3) 给药数据的缺失可能来源于临床试验时的记录错误，应重新核查记录。确认数据的过程中，分析人员应与试验执行者、数据管理者与程序设计师等相关人员共同商讨。

对于缺失用药时间、用药剂量和采样时间的数据，常采用原定的给药方案或者采样时间来替代缺失数据。但该法可能会引入参数估计的偏差。无论采用何种缺失数据的处理方式，均应详细记录于分析报告 (PMAR) 中。此外，若缺失数据的处理方式可造成较大偏差时，须通过敏感性分析评估其对参数估计的影响，并于 PMAR 中呈现。

5.2.2.5 对数转化数据

当数据的范围跨多个数量级时，可考虑对数据进行对数转换，并使用加和型的残差模型或混合型残差模型。对数转换可增加模型与参数估计的稳定性，特别是数值跨度大时效果更佳。诚然，也可视具体

情况，采用原始数据进行建模。

5.2.2.6 非计划数据

非计划数据指未按照研究计划采集的数据、或偏离采样时间窗等方案偏离数据。临床研究中如果收集了非计划数据，经确认数据无误，则非计划数据也须被纳入分析。如果非计划数据不纳入分析，则应在报告中说明原因。非计划数据对模型参数估算的影响可通过敏感性分析予以评估。

5.2.3 探索性数据分析

5.2.3.1. 一般原则

理解数据特征是进行 PK-PD 分析的基础，因此开展群体分析之前，首先应进行探索性数据分析（exploratory data analysis, EDA）。EDA 是一种采用诊断图和统计学方法，探索和揭示数据特征的分析方法。开展 EDA 的意义包括：

- (1) 可视化呈现数据，以了解数据的分布范围和特征，初步判断结构模型和辨识潜在的离群值；
- (2) 确认潜在协变量的分布范围、协变量之间相关性，以避免协变量分析中的混杂效应带来的偏倚；
- (3) 比较不同亚组之间的数据差异；
- (4) 初步验证模型假设是否合理，及时调整模型假设。

基于 PK-PD 分析数据的类型，常用的诊断图参见下表 2。

表 2 常用的诊断图

数据类型	绘图类型	示例
连续或有界评分数据	散点图、折线图、箱线图、直方图、密度图	血药浓度、血压、肿瘤大小、血糖水平、银屑病面积和严重程度（PASI）评分
二分类数据	柱状图、堆积柱状图、ROC 曲线、瀑布图	治疗反应、不良事件发生
分类或有序数据	条形图、堆叠条形图、饼图、累积概率图、马赛克图	疼痛严重程度、不良事件、肿瘤分级
时间-事件数据	Kaplan-Meier 图，累积风险，森林图	总生存期，无进展生存期
计数数据	条形图、点图、泊松分布图	哮喘加重、癫痫发作

统计学方法包括计算变量的中位数、范围、四分位数、算术平均数、几何平均数、标准差等，对数据进行统计学描述。

5.2.3.2. 因变量

群体分析中因变量（dependent variable, DV）通常指血药浓度、生物标志物、有效性与安全性等药效学指标。通过对药物及其主要活性代谢产物的血药浓度、有效性与安全性的药效学指标随时间作图，以及血药浓度对药效指标作图，可加深对药物 PK-PD 特征的理解。

5.2.3.2.1 药动学

PK 数据常为血药浓度，其随时间的变化曲线应采用常规尺度和半对数尺度两种形式作图分析。每个受试者的血药浓度数据点采用连线的方式描述血药浓度随时间的变化趋势。必要时血药浓度随时间的变化曲线可按照亚群体进行分层分析。此外，用于内部建模和外部验证的数据也需作分层分析。

5.2.3.2.2 药效学

PD 数据可为连续变量，如体重、血糖等，在一定范围内取任意值；也可为分类变量，如一定范围内取离散值的不良事件等级，或疾病评估量表分值，或疾病严重程度等；还可为时间-事件数据，如总生存期等。

此外，非连续数据常采用概率或计数模型处理。如果有序分类数据足够多（通常为 6~10 个或以上），则可考虑作为连续数据处理。有序分类数据一般具有明确的数值范围，简单将其视为连续数据进行建模有时并不合理。此时，可考虑结合 Beta 分布，以保证模型化的合理性。此类数据又称为有界评分数据。

基于数据类型，常用的诊断图方法如下：

- (1) 连续或有界评分数据：绘制药效与暴露量的散点图，并添加观测数据的非参数回归曲线。
- (2) 二分类数据：按应答者/无应答者（有效性），或出现不良事件/未出现不良事件（安全性）分层，并对其相应的药物暴露量进行作图（如箱线图）。此外，还可对数据进行暴露量四分位数或三分位数分组，将观察到的反应概率与平均暴露量进行对比，并叠加单变量逻辑（logistic）趋势线。
- (3) 多分类或有序数据：按照事件等级（序数）分层，并对其对应药物暴露量四分法或三分位数作图（条形图或堆叠条形图）。此外，还可针对每个等级（序数），考察其概率和暴露量的关系。
- (4) 时间-事件（time-to-event）数据：按暴露量四分位数或三分位数绘制 Kaplan-Meier 生存曲线。
- (5) 计数数据：绘制计数频率直方图（离散数据），显示数据分布情况。

5.2.3.3. 协变量

协变量包括连续型协变量和分类型协变量。连续型协变量需提供描述性统计结果，并采用直方图或箱线图描述连续型协变量的总体分布等。分类型协变量需提供描述性统计结果，并计算每个协变量各分类的频数和频率。必要时，协变量可按照亚群体进行分层分析，并采用图表形式，探索协变量之间的相关关系。如果纳入多个临床研究进行分析，须比较不同研究之间协变量分布特征的差异。

5.3 基础模型

5.3.1 定义

PPK-PD 的基础模型包括结构模型和随机效应模型。其中，随机效应模型又称为统计学模型。随机效应包括了个体间变异、残差变异、以及其相应的协方差矩阵。

基础模型的假设如药理学假设、生理学假设、疾病假设等应科学、合理。同时，基础模型应能表征体内药物浓度和 PD 指标随时间变化的过程。根据基础模型，可推断药物浓度和 PD 指标与时间的关系、以及随机变异（如个体间变异与残差变异）的大小。基础模型建立之后，方可进行协变量模型的构建。

5.3.2 算法

5.3.2.1 算法特点与选择

本文主要以 NONMEM 软件为例，介绍常用的参数估算方法。NONMEM 7 中的主要算法见表 2，包括一阶估算法（first-order estimation, FO），一阶条件估算法（first-order conditional estimation, FOCE），含交互作用的一阶条件估算法（FOCE with interaction, FOCE-I），迭代二步法（iterative two-stage estimation, ITS），重要性抽样最大期望法（Importance sampling expectation-maximization, IMP），随机逼近期望最大化法（Stochastic approximation expectation-maximization, SAEM），蒙特卡洛马尔科夫链法（Monte Carlo Markov Chain, MCMC），贝叶斯法（Bayesian estimation, BAYES）等。其中，SAEM、IMP 以及处理 BQL 数据的 M3 法均需采用基于二阶近似的拉普拉斯法（LAPLACE）。

表 2 参数的估算方法特点与选择

方法	描述	优点	缺点	适用场景
FO	忽略随机效应和协变量间的相互作用，且根据群体平均值线性化模型。	快速，计算量小	非线性模型准确性略低	低变异的数据，简单线性模型
FOCE/FOCE-I	通过调节个体参数来估算个体间变异。FOCE-I 估算考虑了个体间变异和随机随机变异之间的相互作用。	较 FO 更精确，对于非线性模型也更精确	较 FO 需更大的计算量	较为通用，适用于中等复杂程度的非线性模型
LAPLACE	与 FOCE 相似，使用二阶泰勒展开来提高精度的近似方法	较 FOCE 具有更高的准确性	较 FOCE 需更大的计算量	需高准确性的非线性模型，并可用于离散型或分类性数据（概率模型）的拟合
ITS	二阶段迭代计算方法：一阶段为个体拟合，二阶段为群体拟合。	能处理复杂的模型以及高变异性	需显著增加算力	个体和群体变异性显著的模型

IMP	采用加权模拟近似似然。	灵活，能够处理复杂模型	需更多的算力，计算结果可能不稳定	用于 FOCE 或 Laplace 不收敛的复杂模型中
SAEM	基于模拟的稳健方法，通过随机抽样来近似似然。	能够处理复杂模型与稀疏数据	需较大算力，但是计算结果更稳定	数据稀疏或不完整的复杂模型，广泛用于 PK-PD 建模(使用时常结合 IMP 算法)
BAYES	结合先验分布与似然估算参数。	能提供完整的参数后验分布	较大算力，且需要先验信息	具先验信息，或完整概率参数描述

其他软件如 Phoenix NLME 中，也使用 FO 算法。对于 FOCE 算法，NLME 又分为 FOCE-ELS 和 FOCE-LB 算法。其中，前者与 NONMEM 中的 FOCE 相同，而后的计算速度更快。而与 SAEM 和 IMP 类似的算法分别为 QRPEM (quasi-random parametric expectation -maximization) 和 MCPPEM (Monte Carlo parametric expectation-maximization)。Monolix 中主要使用 SAEM 算法，S-ADAPT 软件也使用 MCPPEM 算法。

参数估算方法的选择取决于数据的多寡与模型的复杂程度。数据越少则参数估算方法越重要。而对于数据量丰富的情况，不同参数估算方法可提供近似的结果。

5.3.2.2 中间变量的引用

NONMEM 更新 EM 和 BAYES 算法后，推荐引入中间变量 MU，以提高 EM 和 BAYES 算法的计算效率。引入 MU 是一种处理 PPK-PD 模型中个体间变异的标准化方法，使模型参数的个体间变异在对数尺度上表达。MU 不仅可更容易地解释个体间变异对参数的影响，而且也确保了模型结构的一致性。对于呈对数正态分布的参数，如清除率 CL 或分布容积 V，MU 法尤为有用。使用 MU 时应注意：

- (1) 除了残差变异外，尽可能在 THETA 上使用 MU 法。
- (2) 所有 THETA 和 MU 间呈线性关系。
- (3) 在\$PK 或\$PRED 中定义 MU，而不在\$ERROR 中定义。
- (4) 应用 MU 时，不可进行条件限制，即不在 IF...THEN 内定义，并尽可能置于原代码之前。

5.3.3 结构模型的建立

建模的第一步是选择结构模型。通常应根据分析目的、数据类型和特征选择结构模型。同时，模型假设如药理学假设、生理学假设、疾病假设、数据处理假设和统计假设等应科学合理。

药动学结构模型常采用一房室、二房室模型。如药物具复杂药动学行为时，也可构建复杂模型。

药效学结构模型的选择常基于药物的作用机制和 PD 数据的类型。如连续型数据可以考虑直接效应、间接效应模型或效应室模型；二分类数据可以选择 logistic 回归模型等。常见的数据类型及相关模型如表 3 所示。对于离散型或分类型数据，常采用 LAPLACE 算法进行分析。

表 3 连续和非连续数据及其常用 PK-PD 模型

数据类型	常用模型	示例
连续数据	直接效应模型（线性或非线性回归模型）、间接效应模型、效应室模型	血糖、血压、血脂、受体占有率、肿瘤体积
二分类数据	Logistic 模型	有效/无效、客观缓解率（ORR）
分类数据或序数数据	比例优势模型（proportional odds model），马尔科夫链模型（Markov chain model）	疼痛严重程度、不良反应分级
时间事件终点	比例风险回归模型#(proportional hazards model; Cox 回归模型)，参数法生存分析#	生存/死亡、总生存率（OS）、不良事件（AE）发生率
计数数据	Poisson 回归、负二项分布模型(negative binomial distribution model)	不良事件计数、哮喘加重次数
有界评分数	Beta 分布结合直接或间接效应模型等方法	银屑病面积和严重程度（PASI）评

据	分
---	---

PD 模型可能包含基线效应、安慰剂效应、生物标志物及替代终点效应、疗效和安全性效应、疾病进展、耐药性的产生等多重因素。

由于 PK-PD 模型常较 PK 模型更复杂，故应根据建模数据和建模目的，选择合理的建模方法。在 PK-PD 的建模过程中可采用同时拟合 PK 和 PD 参数的方法，也可采用序贯拟合，即先拟合 PK 参数再拟合 PD 参数。建模方法的选择须综合考虑参数的拟合精度与计算耗时之间的平衡。当药物的 PK 行为不受 PD 影响时，建议建模时可固定 PK 参数并在 PK-PD 数据集中包括 PK 数据。由于许多生物大分子药物的 PK 可受到 PD 效应的影响，则应考虑采用 PK 和 PD 同时建模的策略。

尽管 OFV 变化可反映模型拟合的情况，但不推荐仅通过 OFV 的变化确定结构模型。不同的 PD 模型间常不是嵌套模型，故无法通过 OFV 进行直接比较。选择模型时，须结合诊断图等进行综合评判。此外，确定结构模型时，还应检视协方差结果中的相关性矩阵等，避免结构性参数间的强相关性。

5.3.3.1 参数

建议对常见的 PK 参数，如 CL 与 V 进行参数化（如 NONMEM 中采用 PREDPP 模块），以简化后续的参数拟合与协变量建模的过程。建立二房室模型时，应将中央室体积与外周室体积参数化为 V_c 和 V_p ，而不用稳态表观分布容积，以便更好地发现与解释协变量效应。复杂的 PK/PD 模型可用微分方程进行参数化。此外，建议考察不同的参数初值对参数估算值的影响，防止参数估算时的局部最小化。

5.3.3.2 结构模型的协变量

如果已知某些协变量对 PK 或 PD 参数产生较大的影响，则建议在构建结构模型时引入这些协变量。例如，对于主要通过肾脏清除的药物，在结构模型中可考虑将该药的 CL 分为肾脏清除率与非肾清除率，并在肾脏清除率中引入肌酐清除率的影响，从机制上表征药物的消除方式。又如，药物代谢酶的基因多态性可影响药物的清除，则可将此引入结构模型。

结构模型中还可考虑采用异速生长缩放法（allometric scaling）引入体重或体表面积。例如，体重数据跨度较大，或模型中同时包含儿童和成人的数据，分析时可先固定 CL 的体重幂指数为 0.75，V 的体重幂指数为 1.0。之后，对幂指数进行重新估算。若可显著提高模型拟合度且参数估算值合理，则应采用幂指数估算值。此外，生长模型（maturation model）也可用于儿童 PK 参数的缩放。

5.3.4 随机效应模型的建立

5.3.4.1 一般原则

群体法建模的最大特点之一是将预测值和实测值的差异区分为 PK/PD 参数以及药物浓度两层，分别进行估计。PK/PD 参数变异可称为第一类差异（也称个体间差异（ETA）），以解释个体预测 PK/PD 参数和群体预测 PK/PD 参数的差异，其方差为第一类变异（OMEGA）。药物浓度实测值和依据个体预测 PK/PD 参数计算的药物浓度预测值之间的差异为第二类差异，即残差（也称个体内差异，EPS），其方差为第二类变异（SIGMA）。这两类差异都可用数学方程表征，称为随机效应模型。评价随机效应时，还应考虑其随机效应的假设是否成立。建模过程中，个体间变异模型和残差变异模型可逐步完善。

5.3.4.2 个体间变异

建议用指数模型来表示 PK 参数的 ETA，以保证 PK 参数均为正数。ETA 可加到数据支持的任一 PK 参数上。例如，CL 与 V 是最常引入 ETA 的 PK 参数。如果吸收相有足够的参数，则 ETA 也可以引入至吸收相参数（如一级吸收速率常数 k_a ）。除非受试者同时有口服给药和静脉给药的数据，否则口服生物利用度的个体间变异难以估算。

对于 PD 参数的 ETA 可用加和型模型来描述。但由于受试对象对药物的敏感性不同或其疾病状态不同，不同类型受试者的 ETA 可能不同，需分别设置不同的 ETA。

如果 ETA 估算值很小，甚至接近零，则并不代表该参数没有 ETA，而可能表明数据不足以预测 ETA，或者数据量少导致个体间变异的收缩（shrinkage）。此时，如采用最大似然法进行估算，则可将 ETA 固定为零。但是，ETA 固定为零时，期望最大化（Expectation Maximization）算法在 NONMEM 7

中是很低效的。建议将 ETA 固定到极小值维持算法的有效性。

5.3.4.3 OMEGA 矩阵

根据研究目的、研究对象和药物特征，建议在引入结构模型协变量之后，再通过全 OMEGA 矩阵（full omega block）来考察 ETA 之间的相关性。有时，由于数据量不够，难以进行全 OMEGA 矩阵的拟合。但建议考察 CL 与 V 的 ETA 之间的相关性。当参数的 ETA 间存在相关性时，如果采用仅有对角线元素的 omega 结构将忽略 ETA 间的相关性，导致模拟时产生的参数组合不符合实际。在构建协变量模型之后，应重新审视 OMEGA 矩阵，ETA 间的相关性可能会被一些协变量替代。

5.3.4.4 残差变异

PK 数据进行对数转换有利于 EPS 模型的稳定。PK 数据对数转化时，常使用加和型或混合型模型。现行 NONMEM 软件并不支持直接计算个体加权残差（IWRES），可用“THETA”化的“SIGMA”来计算 IWRES。

对 PD 数据的 EPS 进行建模的过程基本与 PK 数据相同。但应确定 PD 指标是否可为负值。例如细胞计数等不可为负值。将此类数据进行对数转换，可确保模拟得到数据无负值。

一般，确定结构模型后再确定 EPS 模型。在优化结构模型时，应同时优化 EPS 模型。当所有的 PK 数据都进行对数转化时，推荐使用加和模型来表征 PK 的 EPS。如果加和型模型难以很好表征 EPS，则应考察其他类型的 EPS 模型。

5.3.4.5 场合间变异

模型中是否纳入场合间变异（Inter-occasion variability, IOV）取决于试验目的和设计。IOV 可在引入协变量之前就纳入模型，并可在整个建模过程中都予以考虑。通常可通过研究设计和数据收集的特点来定义不同的场合。当整合不同临床试验数据之时，尤应注意场合的定义，否则难以解释 IOV 的影响。此外，应将 IOV 对参数的影响和参数的时变性相区分。IOV 是一种随机变异，不应随时间推移而呈现趋势性变化。

5.3.4.6 协方差

协方差计算中最重要的结果是参数的协方差矩阵。观测值越多，协方差估算越准确。协方差估算能够定量表征参数估算的精确性。

协方差运算成功时，应审视变量间的相关系数矩阵。如相关系数接近 1，则说明模型存在过参数化且不稳定。协变量运算成功时也可能是获得了局部最小值，非全局最小值。此时，建议查看模型的条件数（conditional number）。如果条件数过大（例如 >1000），则表明协方差运算结果可能不够准确，建议考虑其他方法计算参数的精确性。条件数的计算见 5.3.4.7 节。

协方差运算失败表明模型可能过参数化，应简化结构模型或者随机效应模型。有时，改变初值可使运算收敛，但常仅获得局部最小值。有时，受限于数据，拟合一直失败。但拟合失败时也能提供一些有用的信息，包括参数拟合值等。在此情况下，参数的精确性还可通过自助法获取，可参见 5.5.3 节。

5.3.4.7 相关系数矩阵的特征值

协方差运算的另一个重要结果是相关系数的特征值。模型的最大特征值除以最小特征值等于条件数。条件数标志了模型的稳定性。当条件数>1000 时，模型很有可能过参数化。通过相关系数矩阵，可追溯其原因。VPC 很难发现过参数化。相反，参数越多，VPC 呈现的结果越好。有时，条件数>1000 的模型也有价值。如量化不确定性的建模，或采用有限数据建立机制性模型。

5.4 最终模型

5.4.1 定义

最终模型是在筛选协变量后，保留了所有具显著协变量的最简模型。最终模型应能准确反映药物的 PK/PD 特征，量化协变量对模型参数的影响，并明确协变量与临床用药方案间的关系。模型中的随机变异通常由 ETA、EPS 与 IOV 描述，且能被引入的协变量所解释。

如果不需要协变量筛选（如分析数据集中不包含协变量），可采用不包括协变量的模型作为最终模

型。应注意含有协变量的最终模型未必比不含协变量的模型更加具有预测能力。

5.4.2 协变量模型

5.4.2.1 一般原则

构建协变量模型前，须综合考虑药物属性、研究对象特征、临床前数据、早期临床数据、以及数据探索性分析结果，明确待考察的协变量。待考察的协变量常包括人口统计学特征、实验室检验结果等，如年龄、性别、种族、体型（体重、身高、BMI、去脂重量等）、肾功能、肝功能、基因型、药物制剂类型、合用药物、以及服药前是否进食等。

在协变量被纳入模型前，应探讨每个待考察协变量的科学意义，包括药理学、生物学、病理学与临床意义等。例如，体重、BMI、去脂体重和身高是高度相关的变量，常用于表征体型的大小。但将这些协变量指标全部纳入模型中不仅没有意义，而且将产生共线性问题。又如，研究群体中有一部分是肾功能损伤患者，则可将反映肾功能的生化指标以连续变量或分类变量形式纳入到模型。

构建协变量模型时，应避免过参数化。过参数化将导致收敛失败与协方差运算失败。当两个或以上的协变量高度相关时，或者协变量分布不均衡时，应合理地选择协变量，避免共线性。此外，亦不可为了便利，将连续变量人为转化为分类变量。

对于二房室模型，除非违反生理学原理，协变量通常应在CL与V_c中纳入。尽管房室间清除率Q与外周隔室V_p较少受协变量的影响，但如有充足证据，亦可考虑纳入，例如体重等。

PD效应可能受到多种因素的影响，如疾病进展、生物反馈系统、耐药性、安慰剂等。为了描述药效随时间的变化，在PD建模过程中须考虑上述因素的影响。

具有明确机理支持、且有临床意义的协变量，可能由于样本量等原因缺乏统计学意义，但可根据研究目的纳入该协变量。

5.4.2.2 缺失协变量

对于不同类型的协变量缺失数据，应根据具体情况采用相应的处理方法，并重点关注缺失值的出现是否随机，尤其应关注缺失值是否在特定人群或场景下频繁出现。可接受的缺失数据比例取决于协变量的性质（例如，连续型或分类型、变化型或稳定型等）、协变量对分析的重要程度等。对于缺失率大于10%的协变量，可考虑不对其进行协变量分析。对于缺失率小于10%的协变量，常用的处理方法包括删除所有含缺失协变量的记录；或对缺失协变量进行填补。常见的缺失协变量填补方法包括：

- (1) 对所有缺失值采用群体的中位数或平均数进行填补，当涉及特定协变量（如体重）时，应根据性别选择不同的中位数或平均数；
- (2) 对于不随时间变化的协变量，可用个体历史数据作为填补值；
- (3) 对于随时间变化的协变量（如体重、生化检验结果、年龄等），可采用前后数据的插值或最近一次的观测值进行填补；
- (4) 根据协变量之间的相关性进行填补。

缺失值处理不当可能导致参数估计偏差，影响研究结论的准确性。在临床试验过程中，建议采取多重措施减少缺失值的发生。此外，缺失值还会增加结果解释的复杂性。在制定分析计划时，应预先考虑缺失值的处理策略，并根据需要进行敏感性分析，以评估缺失值可能造成的影响。

5.4.2.3 协变量的筛选方法

建议采用系统化的协变量筛选方法，以确保协变量模型建立过程的统一性和一致性。选择协变量筛选方式时，应注意不同筛选方法的优势和局限性。此外，协变量筛选前，应确保基础模型的稳定性，并明确协变量的分布特征，避免出现共线性的问题。常用的协变量筛选方法包括以下两种方法：

(1) 全模型估计法（full model estimation, FME）：将所有具潜在影响的协变量同时整合入模型，然后进行一步协变量分析，以构建最终模型。

(2) 逐步法（stepwise covariate modelling, SCM）：SCM包括前向纳入（forward inclusion）和逆向剔除（backward elimination）两个过程。前向纳入过程中，逐一考察和纳入协变量，形成全协变量模型（full covariate model, FCM）。逆向剔除过程是在全协变量模型基础上，逐一剔除协变量，形成最终模型。

SCM 筛选协变量时，使用全协变量模型有一定优势。例如，协方差矩阵可提供有用的模型诊断信息，还可评估协变量极值相关的离群值。此外，全协变量模型还可提供协变量对模型参数的影响程度。但是，由于 SCM 中的前向纳入过程是逐一考察协变量，故难以对特定组合的协变量效应进行评估。

相较于 SCM 法，FME 法侧重于模型参数估算值的合理性和准确性（包括相对标准误和自助法的置信区间），可直接评估协变量的临床意义，而不依赖于目标函数值（OFV）的变化等数据驱动的评判标准。FME 法只需运行一次即可完成协变量的评估，结果易于解读。

此外，尚有其他协变量筛选方法，包括瓦尔德似然比检验近似法[Wald's approximation to the Likelihood Ratio Test (LRT), WAM]、广义加和模型（generalized additive modelling, GAM）法、拉索法（least absolute shrinkage and selection operator, LASSO）法等，不推荐常规使用。附件表 A 展示了各类筛选协变量方法的特点，供参考。

5.4.2.4 协变量模型的构建

协变量模型应能描述数据的整体趋势，准确表征协变量的影响。建立协变量模型的过程中，可使用加和型、比例型、指数型、幂函数型等数学关系式，确保参数估算值的合理性，如 PK 参数值应为正数。

协变量模型构建时可采用中心化方式。中心化的协变量能不影响参数估算的情况下，有利于参数估算结果的解读，并提高参数估算的稳定性。一般而言，协变量的中位数（连续型）或者众数（分级型）的数值可作为中心化的参考值。在协变量的数值跨度很大（如儿童与老年人的混合群体）时，应综合考虑数据特征和分析目标，确定中心化的参考值。

5.4.2.5 协变量参数的边界

一般而言，不对协变量参数设定边界值，而可用特定方式进行数据转化，如协变量的对数转化可使协变量参数值的更合理。有时，可设定边界来限定参数估算的范围，减少计算时间，避免产生不合理的参数值。

5.4.2.6 协变量模型的充分性

使用全协变量模型构建协变量模型时，需确认全协变量模型是所有模型中 OFV 最低的模型。由于全协变量模型中纳入了最多的协变量，因此如果全协变量模型的 OFV 比基础模型高，则说明全协变量模型在运算中仅获得了局部最小值。

5.4.2.7 结果解读

协变量模型成功建立时，应对最终模型估计值进行正确解读。

(1) 如果置信区间很窄，并且包含 0 值，则数据不支持这个协变量效应

(2) 如果置信区间很窄，并且不包含 0 值，则数据支持协变量效应

(3) 如果置信区间很宽，但不包含 0 值，则可能有协变量效应，但分析用数据难以准确估计效应大小。

上文中的“窄”和“宽”应基于临床相关性与重要性进行定义。此外，无论协变量的影响程度有多大，均应说明协变量的临床意义以及相关原因。分析结果可用森林图（forest plot）等展示协变量对 PK 参数或暴露量参数（如 C_{max} 和 AUC）的影响。

5.4.3 随机效应模型

加入协变量模型后，随机效应模型可能需要优化，其构建方法与基础模型部分相似，参见 5.3.4 节。

5.4.4 其他

5.4.4.1 离群值

离群值指在同一个受试者中明显偏离其他观测值的异常观测结果，或是明显偏离其他受试者的观测结果。建议将 CWRES 绝对值大于 5 或 6 的数据定义为离群值。在建模过程中，包括基础模型、全协变量模型与最终模型的构建过程中均应对离群值予以关注。一般而言，离群值不应被剔除，但有时离群值可影响模型拟合的成功，应予以剔除。剔除离群值后，应开展敏感性分析，并将结果记录于分析报告。

5.4.4.2 过参数化

在 PPK-PD 模型建立过程中，可遇到最小化失败的情况。而过参数化是导致最小化失败的常见原

因。根据以下信息，可采取相应的解决方法。具体如下：

(1) 梯度 (gradient)：在最小化过程中出现以下的梯度值是非正常情况，说明数据信息不足以支持某个参数的估算，或编程错误，或数据缺失。

- 初始迭代或最后一次迭代时，梯度为零
- 梯度在极大值与极小值之间波动，例如从 $1.0E+07$ 到 $-1.0E+07$
- 梯度较初值的变化很小，例如从 $1.0E+07$ 到 $1.0E+06$

(2) 初值：在最小化过程遭遇失败时，也可考虑采用重新设定初始条件的方法，探索参数的估计值。若识别出问题参数，通过调整这些参数的初始值，可促进最小化过程的成功。在 NONMEM 软件应用中，相较于 OMEGA 参数，调整 THETA 参数的初始值通常更为有效。

(3) 协变量：如果缺乏协变量与参数的关系，则改变协变量与参数间的数学关系不会解决模型的不稳定性。

(4) 边界值：若由于撞界致最小化终止，应检查梯度为零的参数，然后移除该参数或放宽参数的边界值。

(5) 舍入误差：如果最小化终止是由于舍入误差 (rounding errors) 引起，则可改变 NONMEM 输出的最低有效数字的个数 (number of significant digits, NSIG)。例如，最小的 $NSIG \geq 2$ 时，则可在 \$EST 中设置 $NSIG=2$ 。如果最小的 $NSIG < 2$ 时，则模型可能过参数化。此时，可考虑将 NSIG 低的参数固定或简化模型。

(6) 目标函数值趋向无穷：若最小化错误源于目标函数值趋向无穷大，则缺乏统一的解决方案，需根据具体情况进行判断。

5.4.4.3 基于先验信息的模型

在 NONMEM 7 及更高版本中，引入了一种基于先验信息的贝叶斯建模方法，即 PRIOR 法。该方法通过整合先验信息，改善模型的稳定性和预测能力。数据量相对有限时，PRIOR 法能充分利用先前的研究结果进行模型构建。采用 PRIOR 法时，应综合考虑以下几个方面：

- (1) 数据的质量：虽然 PRIOR 法可以在数据量有限的情况下改善模型的稳定性，但高质量的数据仍然是构建准确模型的关键。
- (2) 先验信息的合理性：先验信息应基于可靠的来源，如既往研究结果或专家共识等。不合理的先验信息可能会导致模型的偏倚。
- (3) 计算量：引入先验信息会增加模型的计算量，常需要更多的计算资源和时间。

此外，先验信息也可以通过固定参数值的方式引入模型，即将参数固定为特定值。无论使用何种方法引入先验信息，建议在可行的情况下进行敏感性分析，并对方法选择的原因和合理性进行解释、并记录于分析报告中。

5.4.4.4 混合模型

当总体研究对象由若干个具有不同特征的亚群体 (又称为“混合群体”) 构成时，可应用混合模型描述不同亚群体的 PK/PD 行为。例如，患者人群中具不同的清除率或吸收特性的亚群体，可采用混合模型进行表征。

在 NONMEM 中，可通过为每个亚群体指定不同的特征参数来构建混合模型，估算个体属于每个亚群体的概率。例如：药物代谢酶具有基因多态性，导致两个代谢亚群体：快代谢者和慢代谢者。混合模型可估计两个亚群体的清除率以及每个亚群中的个体比例。此外，构建混合模型时，应注意如下几个方面：

- (1) 模型复杂性：混合模型提升了模型的复杂程度，需额外的参数描述每个亚群体及其隶属的概率。
- (2) 参数可辨识性：区分亚群体时，可能由于参数值差异很小或者数据量不足估，可导致难以区分“亚群体”，而无法估算参数值，即参数不可辨识。
- (3) 计算收敛：由于模型的复杂程度增加，混合模型可能难以计算收敛。数据有限或存在噪声的情况时，亦可能难以找到一个稳定的收敛解。
- (4) 过拟合：模型过度适应特定数据集 (即过拟合)，可丧失了泛化到新数据的能力。因此，选择合适的模型，并进行有效的验证是避免过拟合的关键步骤。
- (5) 模型评价：进行混合模型的评价时，需注意自助法中亚群的比例以及视觉预测检查 (Visual Predictive

Check, VPC) 的合理性。

(6) 结果解释: 混合模型在解释临床亚群体形成的原因方面常面临困难, 故应用时常受到限制。

5.5 模型评价

5.5.1 定义

模型评价指客观评价模型在目标应用领域中的预测能力, 或判断模型偏倚是否会对决策产生实质性影响。模型无法普适于所有场景, 但可评估模型是否适用于特定场景。模型评价是群体分析的重点内容之一, 贯穿于模型构建与应用的全过程, 涵盖了对模型结构、模型参数和模型预测性能等多方面的评估。应根据研究目的选择模型评价的内容以及评价标准。如果评价结果不符合研究目的时, 应对模型进行优化和调整。

通常, 模型评价通常可分为以下几类:

- (1) 基于预测值的模型评价: 如群体预测值 (population prediction, PRED), 个体预测值 (individual prediction, IPRED);
- (2) 基于残差的模型评价: 如权重残差 (weighted residuals, WRES), 条件权重残差 (conditional weighted residuals, CWRES) 等;
- (3) 基于经验贝叶斯估计 (empirical Bayesian estimation, EBE) 的模型评价;
- (4) 基于模拟的模型评价。

5.5.2 一般原则

通常, PK模型的评价方法与采用连续数据的PD模型的方法一致。但在处理非连续PD数据时, 由于PD模型在拟合过程中无法生成预测结果, 基于预测的诊断图可能不适用。在这种情况下, 通常采用基于模型模拟的方法来进行模型评价。例如, 评估肿瘤患者的生存期时, 可通过对比实际观测获得的Kaplan-Meier曲线与模型模拟的曲线来进行模型评估。

由于非连续PD数据涵盖多种类型, 因此具体的评价方法也各有差异。应采用能反映其数据特性的评价方法, 对比观测值与模型模拟值。在评价非连续数据的模型时, 通常只能提供基于群体或亚群体层面的评价结果, 而难以提供针对个体的评价结果。

在评估PD模型时, 药物的药理特性也应作为考量因素。例如, 对于使用国际标准化比值 (INR) 作为PD指标的抗凝药物模型, 须能表征INR的最小值和最大值; 对于部分抑制作用的模型, 应验证其最大效应是否能超100%; 而对于降糖药物的模型, 需检查其最大效应是否与真实的生理状况相吻合。

由于PD数据通常比PK数据更为复杂且多变, 如肿瘤生长数据往往呈现更大的异质性和非线性特征。因此, PD模型的诊断图往往难以呈现出一致的趋势性, 基于预测值的诊断图可能表现不佳。PD参数也往往不像PK参数呈现已知的分布趋势。在这种情况下, 应当谨慎解读PD模型的评价结果。

5.5.3 诊断图评价

5.5.3.1 一般原则

诊断图的应用过程中须关注以下问题:

- (1) 拟合不佳时, 应注意结构模型是否合适;
- (2) 是否违反非线性混合模型中随机变异分布的假设, 包括: 多峰、不对称、均值不为零或方差不齐等;
- (3) 离群值;

绘制诊断图时, 可根据以下因素开展分层分析, 便于鉴别拟合不佳或识别模型的影响因素:

- (1) 实验设计: 实验中含不同研究组或不同治疗组时, 可按照不同组别进行分层分析。
- (2) 协变量: 根据是否纳入协变量进行分层分析、或者根据连续变量数值大小分为若干组进行分层分析, 以检视协变量的影响。
- (3) 时间: 如果研究周期很长, 可根据时间分层, 以审视不同时期的影响

在建模过程中, 建议对于重要的模型开展以下诊断图评价:

- (1) 应用普通尺度与对数尺度, 绘制观测值-群体预测值 (OBS-PRED) 散点图, 并加上趋势线与参考

线;

- (2) 应用普通尺度与对数尺度, 绘制观测值-个体预测值 (OBS-IPRED) 散点图, 并加上趋势线与参考线;
- (3) WRES 或 CWRES 对 PRED 的散点图;
- (4) WRES 或 CWRES 对 TIME 或者给药后时间 (time after dose, TAD) 的散点图;
- (5) OBS、IPRED 或者 PRED 对时间的散点图 (相互叠加或者并排放置);
- (6) IWRES 绝对值对 IPRED 的散点图;
- (7) IWRES、WRES 和 CWRES 的频数图或者 QQ 图;
- (8) 视觉预测检验 (visual predictive checks, VPC) 或预测值校正的视觉预测检验 (prediction-corrected VPC, pcVPC)

对于基础模型和最终模型均需开展上述模型评价, 并可并排展示基础模型和最终模型的诊断图, 以展示协变量纳入之后的改进。最终模型还应包括以下拟合优度图:

- (1) 经典贝叶斯参数估计值 (EBE) 与对应 ETA 的散点图
- (2) 经典贝叶斯参数估计值 (EBE) 和对应 ETA 的分布图
- (3) 经典贝叶斯估算的个体 ETA 值对协变量作图
- (4) 个体拟合曲线

在使用经典贝叶斯参数估计值 (EBE) 进行诊断图评估时, 须注意 ETA 和 EPS 的收缩效应。同时, 应注意模型诊断图对删失数据或缺失数据未作处理。因此, 在分析包含大量删失或缺失数据时, 应谨慎解读诊断图所呈现的结果。

5.5.3.2 基于预测的诊断图

5.5.3.2.1 观测值-群体预测值散点图

观测值-群体预测值 (OBS-PRED) 散点图可评估 PRED 能否准确描述数据的总体趋势。若 PRED 显示系统偏差, 则模型难以准确描述数据的整体趋势。此时应考虑对结构模型或者统计学模型进行修正。此外, 将协变量加入前后的 OBS-PRED 散点图并列展示, 可直观地检视协变量的影响。

该诊断图的局限性在于缺乏一个理想的参照图。除了建模错误之外, PRED 在对角线两侧的分布还受到多种因素的影响, 包括参数变异的程度和数据的删失情况等。此外, 趋势线未涉及预测误差的方差齐性问题, 也未考虑数据可能源自不同的个体。

5.5.3.2.2 观测值-个体预测值散点图

观测值-个体预测值 (OBS-IPRED) 散点图旨在评估 IPRED 是否能准确描述个体观测值。IPRED 在对角线两侧的分布仅揭示了残差变异的大小。若要据此识别模型中的问题, 必须确保每个个体的数据都含有足够的信息以准确估计个体参数。然而, 若个体数据量不足, 可能会导致模型过度拟合。采用此类诊断图作为评判依据前, ETA 的收缩一般应 < 30%。此外, IPRED-OBS 散点图仅反映 EPS 的大小, 而协变量是用来解释 ETA 的来源。因此, 无论是在基础模型还是在最终模型, IPRED-OBS 散点图是相似的, 主要用于评估基础模型的拟合情况。

5.5.3.2.3 个体药时曲线

个体药时曲线对于评价个体的拟合程度十分重要。PRED 和 IPRED 的药时曲线可重叠呈现、或者并排呈现。

5.5.3.3 基于残差的诊断图

基于残差的诊断图能对预测误差进行直观的评价。不同的残差诊断图能显示模型在结构或者统计学的不同特性。

5.5.3.3.1 加权残差—时间散点图

加权残差 (WRES) 对自变量的散点图可用来判断结构模型的合理性。对于 PK/PD 模型, 时间是关键自变量, 可以用药后时间 (time after dose, TAD) 或者是试验中一段持续的时间。如果时间跨度很

大,则可将时间数转化后作图。如果采样密集,那么可以针对不同的给药时间段作图。例如首次给药后或者是末次给药后的时间。

基于 FO 算法估算的 WRES 值有时不准确,可导致错误地提示模型误设(model misspecification)。而 CWRES 是直接基于 FOCE 算法估算而来,故一般应采用 CWRES 来代替 WRES。

5.5.3.3.2 加权残差—群体预测值散点图

CWRES/CWRESI 对 PRED 散点图优于 WRES-PRED 散点图。散点图中应有水平参考线与 CWRES/CWRESI 的平滑线。如果 PRED 的数值范围很广,则应考虑对 PRED 进行了对数转化后作图。

|IWRES| vs. IPRED 散点图常用于验证残差模型中的假设,包括均数为 0、以及方差齐性等。理想情况下,|IWRES|应在整个 IPRED 范围内呈水平状。

5.5.3.3.3 加权残差的直方图与 Q-Q 图

CWRES/IWRES 的直方图可描述残差的分布。残差分布应为单峰、且均值为零的对称分布。当使用 CWRES /IWRES 直方图评估个体内随机效应的分布时,应注意 CWRES 直方图中不对称的现象可能会更严重。此外,应说明采用核心密度光滑线(kernel density smooth line)还是正常密度估计方法(normal density approximation)。根据观测值数量,可对直方图进行适当分组。另外,IWRES 可通过 QQ 图进行正态分布审视。ETA 收缩值较大时,IWRES 会更偏离正态分布。

5.5.3.4 基于经验贝叶斯估计的诊断图

经验贝叶斯估计 (EBE) 的参数评价依赖于相应的参数 ETA。个体参数的估算值是通过贝叶斯方法综合先验群体参数、残差分布和个体观测值计算而来。

5.5.3.4.1 个体间变异和残差变异

EBE 向群体均值趋近的现象被称为 ETA 收缩。同样,当数据量不充分时,IWRES 的分布向零趋近的现象被称为 EPS 收缩。当具临床意义的关键参数(如清除率)的 EBE 收缩超过 30%时,基于 EBE 的模型评价结果(如 IPRED)可能不可靠。EBE 收缩还可影响协变量和暴露-效应关系的分析,因此须结合其他模型评估方法,综合评估收缩的影响。在对不同研究设计的试验数据进行分析时,有时仅有一个数据子集包含了特定参数的丰富信息。此时,应根据研究设计对数据进行分层分析,并计算各分组的收缩。

5.5.3.4.2 个体间变异和参数的散点图

分析过程中,可采用散点图考察参数间的相关性,并在协变量模型的建立过程中,对参数相关性予以重视。

5.5.3.4.3 个体间变异的直方图与 Q-Q 图

ETA 的直方图和 QQ 图可检验参数的 ETA 是否呈正态分布,以及 ETA 模型是否合理。

5.5.3.4.4 个体间变异与协变量散点图

ETA 与协变量散点图用于审视参数与协变量之间的相关性。其展示方式应依据变量的性质(连续型或分级型)来决定。连续型协变量适宜采用散点图,而分级型协变量则适合使用箱型图。这种诊断图旨在探索而非确立相关关系。建议在构建协变量模型的过程中绘制该图。

5.5.3.4.5 其他

在 NONMEM 计算参数时,采用统计学检验 ETA 的均值是否等于零。针对重要的 PK 参数,如清除率和中央室分布容积等,如果 ETA 的均值不等于零,且 p 值具有统计学显著意义,则应进一步考察其原因,必要时优化或修正模型。

5.5.3.5 基于模拟的诊断图

5.5.3.5.1 定义

基于模拟的模型诊断是通过构建的模型生成模拟数据,评估模型的拟合优度。这种方法有助于确定模型是否能准确描述观测数据,并评估其预测能力。基于模拟的模型诊断具有多种形式,能够解决不同的问题。在执行时,需明确通过模拟解决的问题,以便选择恰当的评估方法。常用方法包括视觉预测检验(Visual Predictive Check, VPC)、数值预测检验(Numerical predictive checks, NPC)、正态预测分布误差(normalized prediction distribution errors, NPDE)、后验预测检验(Posterior Predictive Check, PPC)等。

5.5.3.5.2 视觉预测检验

5.5.3.5.2.1 定义

VPC 是通过对比观测数据与模拟数据的中位数和特定百分位数，来评估现有模型是否准确表征了建模数据的集中趋势（central tendency）和变异程度。VPC 过程中的模拟采用了最终模型参数估算值，不考虑参数的不确定性。此外，VPC 使用观测数据的输入值（给药记录、时间、协变量等）来生成个体模拟数据，无法预测新的应用场景结果。

5.5.3.5.2.2 一般原则

进行 VPC 时，应遵循以下一般原则：

- (1) 模拟次数：VPC 的模拟次数取决于评价目的，例如主要评估数据点集中趋势（central tendency）还是数据的分布是否有拖尾等。大多数情况下，只需模拟 1000 次。如果需要更高的预测准确度，则需要更多的模拟次数。例如，当 C_{\max} 与安全性相关， C_{\min} 与有效性相关，须重点关注时，则需要更高的预测准确度和模拟次数。
- (2) 百分位数：使用哪个百分位数来呈现 VPC，取决于每个分组内的观测值数量。有时，模型的不同部分（结构模型、误差模型等）可能需要不同的百分位数，以保证每个分组内的数据量。例如，每个分组观测值为 10-20 个，可选择 80%（10%-90%）区间，每个分组观测值 >100 时，可选择 90%（5%-95%）区间，每个组内观测值 >200 时，可选择 95%（2.5%-97.5%）的区间。
- (3) 分组：在计算分位数的时候，需对自变量进行分组。最常见的情况是根据时间进行分组，通常将时间相近的观测值合并在同一组内。例如，如果将给药后 3.9 小时后采样与 4.1 小时后的采样观测值放在一个组内。但对于口服药物的吸收相，可能不宜将 0~1 小时内观测值归在同一组内。当稀疏采样时且仅有谷浓度数据时，可采用首次剂量后时间、体重等协变量，开展 VPC。
- (4) 分层：应对重要协变量进行 VPC 分层分析，还可对重要的试验设计特征（例如剂量、试验人群等）进行分层分析，考察上述因素对模型拟合的影响。对于确证性研究，须对重要协变量进行分层考察，在此过程中，应注意不同分层中分组内的观测值数量。
- (5) pcVPC：如果根据多个协变量分层，可能导致 VPC 的诊断效力下降。此时，可采用 pcVPC。pcVPC 使用分组内中心时间点处对应的数值来标准化观测值与模拟值。在模型中包含协变量的时候，建议使用 pcVPC 来诊断。注意 pcVPC 中 y 轴的数值不是实际的观测值。
- (6) BLQ 数据：当存在 BLQ 数据的情况下，应对高于 LOQ 的连续数据观测值和 BLQ 数据分别进行 VPC。

此外，应注意上述因素是相互关联的，开展 VPC 时应进行综合考虑。其选择的依据须在分析计划和报告中明确阐述。

5.5.3.5.2.3 结果解读

VPC 的重点考察模拟数据与观测数据之间相似性，在结果解读时应注意以下问题：

- (1) (100-n)%的观测值在预测区间之外；例如 5%的点在 2.5%-97.5%的百分位数区间之外，其中分别有 2.5%的点在 2.5%和 97.5%预测区间之外
- (2) 在 n%区间之外的点应随机分布，无聚集现象。
- (3) 如果模型的结构合理，那么每组内的模拟值与观测值的均值/中位数是相近的。
- (4) VPC 难以判断模型的误设。当缺失数据多或者采用分级数据时，可考虑使用其他形式的 VPC。

5.5.3.5.3 其他

除了 VPC 以外，还有其他基于模拟的评价方法，例如 NPC、NPDE、PPC 等。

NPC 的原理与 VPC 相似，但侧重于数值统计量的比较。NPC 通过基于待评价模型的模拟数据，构建多个预测区间（如 0、20%、40%、50%、60%、80%、90%、95%），然后统计观测数据在预测区间以外的数量和百分比，并与预期值作比较。

NPDE 是基于整体预测分布评价非线性混合效应模型的标准和检验方法。该法通过对预测差异（prediction discrepancy）进行标准化转换，考虑了同一个体内观测值的影响。若模型有效，则 NPDE 应服从标准正态分布。

PPC 也是一种基于模拟的模型评价技术。该法首先从模拟数据获取 PK/PD 参数的后验预测分布，然后与原始数据进行比较。通常选取多个参数对模型的预测性进行评估。例如 C_{\max} 、 C_{\min} 、 t_{\max} 、AUC、 EC_{50} 或 E_{\max} 等。PPC 定义贝叶斯 P 值为后验预测分布（模拟）统计量 \geq 观测统计量的概率，以评估两者的相符程度。贝叶斯 P 值接近 0 或 1 提示模型可能无效。通过绘制直方图，可直观地呈现模拟数据与

观测数据的分布特征，也可计算理论值与观测平均值发生偏倚的概率，评估模型的准确性。

5.5.4 数值统计评价

5.5.4.1 定义

数值统计评价适用于从基础模型到最终模型的整个建模过程，可评估模型的拟合程度、参数估计的稳定性，以及是否违反模型假设等。常用的数值统计评价包括目标函数值、参数估算值及其标准误、收缩值和自助法等。

5.5.4.2 目标函数值

目标函数值 (OFV) 代表模型与数据的拟合程度。针对同一数据集，OFV 越小表明拟合度越佳。此外，基于最大似然 (Maximum Likelihood, ML) 理论，嵌套模型间的 OFV 变化符合卡方分布，故 OFV 可用于嵌套模型的比较。OFV 变化 >3.84 (自由度=1, 自由度=模型间参数数量的差异)，表明 OFV 较低的模型在统计上更优 ($p<0.05$)。但是，该假设依赖于算法 (如 NONMEM)，可存在较大的 I 类错误。因此，仅依赖 OFV 可能会产生误导，应避免采用 p 值评价嵌套模型间的差异，须结合多种评价方法进行综合评价。

对于非嵌套模型，可用赤池准则 (Akaike Information Criteria, AIC) 对模型进行评估：

$$AIC = OFV + 2 \cdot p,$$

式中 p 代表模型的参数数量。AIC 较小的模型被视为是较好的模型。尽管 OFV 对模型选择的意义有限，但有时仍可提供有效信息。例如，纳入某个参数能使模型能使 OFV 降低 10 以上，则此参数可提高模型的拟合程度。如果有充分生理或生物学证据支持该参数的纳入，则纳入标准可适当放宽。

5.5.4.3 参数估算值及其标准误

评估参数估值是否合理，并且具临床意义的关键参数的相对标准误 $<30\%$ ，非关键性参数的相对标准误 $<50\%$ 。过高的标准误表明参数不可辨识或模型过拟合，也可导致参数估计值的不稳定性。

5.5.4.4 个体间变异的收缩

收缩表征了 ETA 趋近于群体均值的程度。 $>30\%$ 的收缩表明数据未含足够信息准确估计 ETA，导致个体预测值存在潜在偏差。

5.5.4.5 自助法

自助法 (bootstrap) 采用重采样技术，从原始数据中生成多个数据集，评估参数估计的稳定性和变异性。自助法可计算参数估算值的置信区间，评估参数的精确度和稳健性。自助法无须对参数分布进行假设，即可获得参数的经验分布。有时，为了获得合理的参数分布，可根据重要协变量进行分层 (如剂量、性别等)，生成自助法数据集。但是，自助法计算量大，难以辨识复杂模型中不确定性的来源。一般，自助法每个参数的 $2.5\% \sim 97.5\%$ 区间估算结果应包含原始模型的参数估算值，另应满足预先设定的稳健率，如稳健率 $>80\%$ 。

当自助法应用受限时 (如运行时间过长)，可通过抽样重要性重抽样法 (Sampling Importance Resampling, SIR) 估算参数的精确性。在 SIR 法中，可将 NONMEM 估算的协方差初步计算结果或少量次数的自助法结果作为初始点。如不输入初始点，可大幅增加 SIR 法的计算时间。

5.6 模型应用

5.6.1 估算个体 PK 参数和暴露量

PopPK 模型可用于估算受试者的个体 PK 参数 (如清除率、分布容积等) 支持后续的 PopPD 分析和药品说明书中临床药理部分的撰写，也可用于估算受试者的个体暴露量，如 C_{max} 、 C_{min} 和 AUC 等，支持后续的暴露量—效应 (E-R) 分析。暴露量估算时应注意：(1) 用药方案的终止或改变；(2) 由于内在或外在因素 (如疾病状态或抗药抗体等) 的影响，暴露量随时间的变化。

PopPK 模型能够预测个体在不同给药方案下的药物暴露量，例如稳态谷浓度等，从而为个体化用药提供指导。PopPK 模型还可根据协变量预测体内的 PK 过程。这些协变量包括不随时间变化的协变量 (如性别、遗传多态性等) 和随时间变化的协变量 (如体重、肝肾功能、抗药抗体等)。

5.6.2 考察协变量的影响

PopPK/PD 模型能够阐释内在和外在因素引起的 PK/PD 变异，并评估协变量对 PK/PD 的影响。例如，通过固定效应及其参数的估算精度，可构建反映协变量对 PK 参数（例如 CL、AUC、 C_{\max} 和 C_{\min} ）和 PD 参数（如 E_{\max} 和 EC_{50} ）影响的森林图，有助于解释协变量对 PK/PD 的影响和临床相关性。

5.6.3 模拟

5.6.3.1 一般原则

模拟作为群体模型的关键应用之一，能在避免直接进行数学求解的情况下，预测药物的暴露量和效应，评估治疗方案的有效性，从而为临床合理用药提供支持，或有助于研发和监管的决策。若模型能提供 PK-PD 参数，并阐释药物与机体之间的相互作用，则能进行不同条件下的假设性分析，更系统、全面地审视药物、机体与疾病之间的相互关系。

开展模拟前应制定详细的计划，确保模拟工作的顺利进行和对模拟工作的充分认识。模拟过程中应注意以下内容：

- (1) 模拟的患者特征与真实患者的典型特征相匹配；
- (2) 模拟场景在临床实践中具有可行性；
- (3) 研究设计、样本量和研究终点须合理；
- (4) PK 或 PD 指标（AUC、 C_{\max} 、 C_{\min} 、 E_{\max} 和 EC_{50} 等）须合理；
- (5) 根据模拟目的，所采用的模型可以是包含变异的群体模型，也可以是仅呈现群体平均水平、而不考虑变异的群体模型。

应依据在不同假设条件下模拟结果的分析，制定相应的决策。多学科团队对模拟结果的解读，能显著提升决策的制定和完善。此外，研究结果应尽可能通过图表形式展现和总结，以便于具有不同专业背景的专业人员理解，并为决策提供更丰富、更直观的量化支持。

5.6.3.2 类型

依据分析目标，利用多维度信息，执行模拟分析，涵盖但不限于固定效应、参数的不确定性、个体间变异及其相关性、场间变异、随机残差等方面。

5.6.3.2.1 基于固定效应估计的模拟

若模拟分析旨在呈现典型受试者药物暴露或 PD 指标随时间变化的趋势，则仅需依据 PopPK-PD 模型的固定效应参数估计值进行模拟。然而，基于这些典型参数的预测并不代表受试者群体的平均预测值。平均预测值是通过整合模型的固定效应、个体间变异以及随机残差，进而模拟得出的平均药物浓度或 PD 指标随时间变化的曲线。

5.6.3.2.2 基于固定效应估计不确定性的模拟

若模拟分析旨在揭示典型受试者药物暴露量或 PD 指标达到或维持在特定阈值的概率，或旨在阐释协变量对模型参数的作用，须将参数估计的不确定性纳入考量。例如，针对 PopPK 模型固定效应及其参数不确定性的模拟分析，能构建反映协变量对 AUC 或其他暴露量影响的森林图，进一步理解协变量对药物暴露的影响。此类分析对于评估新给药方案在后续临床试验中的表现也具有重要价值。

5.6.3.2.3 基于个体间变异的模拟

在进行模拟分析以展示个体药物浓度或 PD 指标的预测范围时，须考虑模型的固定效应和个体间变异。若分析的目的是进一步揭示个体药物浓度或 PD 指标的实际观测值的范围，则需额外考虑模型的残差变异。若目标是预测未来人群的观察值范围，还应考虑个体间变异的不确定性。

此外，在模拟分析中，应考虑随机效应之间的相关性，以避免受试者中出现与实际情况不符的参数组合。对于需要考虑个体间变异和协变量效应的模拟分析，应在具有真实人口学特征的模拟群体中进行。人口统计学变量可从相关数据库中获取、对原始研究中的受试者个体进行重抽样，或者通过目标人群中的协变量分布及其相关性来生成。

5.7 分析报告

分析报告通常包括九大部分：摘要、研究背景、目的、数据、分析方法与假设、结果、讨论、结论、

附录。 以下对各部分进行描述:

5.7.1 摘要

摘要作为报告的核心,必须以简洁明了的语言呈现,涵盖研究的背景、目的、主要假设、研究设计、数据与方法概览,以及关键结果和结论。此外,对于决策具有显著影响的图表也应包括在内。

5.7.2 研究背景

围绕研究目标,详细描述待分析药物的特性,探讨研究背景以及进行群体分析的意义。

5.7.3 目的

明确研究的目的,包括主要和次要目的。如与方案不一致,则需解释说明。常见目的包括:分析协变量对 PK-PD 的影响,通过模拟选择合理的用药方案等。

明确研究目标,涵盖主要目标和次要目标。若研究方案与这些目标存在偏差,则必须提供相应的解释。常见的研究目标包括:分析协变量对 PK-PD 关系的影响,以及通过模拟确定合理的用药方案等。

5.7.4 数据

详细阐明用于群体分析的临床试验信息,包括研究设计、样本量(例如受试者数量和血药浓度样本量)、采样方案、受试者状况、给药详情(如药物名称、剂量、给药间隔、用药依从性)、药物浓度检测方法以及定量下限、药效评估方法。提供人口学统计描述,并利用图表来探索协变量的特征及其相关性。阐述衍生变量的计算方法、数据格式、质量控制流程以及数据整理程序。若存在不同的分析数据集,需明确描述它们之间的差异,尤其是不被纳入分析的数据及其排除的原因。对于原始观测数据的任何形式转换,必须详细解释并提供充分的依据。

5.7.5 分析方法与假设

为确保分析结果的可重现性,应提供详尽的细节。建议首先概述分析流程和分析内容,提供总体的分析框架。然后,应详细描述处理各类缺失值、低于定量下限的数据以及异常值的策略和方法。在模型拟合方面,需阐述所使用的算法(例如贝叶斯、非参数最大似然估计、FOCE),涉及的假设(如参数分布),模型选择的标准,以及所用软件及其版本。结构模型应通过示意图和公式详细展示,同时包括协变量模型、协变量筛选方法和标准(如 OFV、临床相关性)、个体间变异、残差变异以及时间相关的随机变异模型。模型评价应具体说明所用的诊断图法或统计方法,包括敏感性分析等。对于复杂的模型化分析,应详细说明分析流程。模拟方案的描述应详尽,包括虚拟人群的生成方法等。在分析报告中,须重点阐述显著影响研究结论的假设,并对假设的任何改动进行详细说明和评价。

5.7.6 结果

通常涵盖结构模型、协变量筛选过程中的关键步骤及其评价,以及最终模型、参数估计值和变异性(CV%)、精确度(RSE%、95% CI)等。最终模型的评估是重点,包括模拟、关键协变量对 PK 的影响、剂量调整等。分析结果通常以图表形式展现,图表中的符号需明确标注,以便于理解,有时可采用彩色图表以区分不同指标。结果应集中于研究目标和模型应用,其他内容可置于附录中。

5.7.7 讨论

讨论不应是结果的重复,而应深入解释建模结果,并在现有研究的基础上阐述这些结果的临床意义。围绕研究目标,评估模型在解决问题方面的效能,包括用于建模的数据是否充分及其潜在局限性;建模方法和假设验证的原理,对模型结果(例如协变量及其对暴露的影响)的阐释,以及其它研究如何支持这些结果(例如类似药物的研究)。此外,讨论还应涉及模型结果与传统单个临床药理研究之间的差异。模型结果对于指导给药方案的选择也非常重要,例如体重和肌酐清除率对药物清除的影响,以及针对特殊人群的给药方案调整等。

5.7.8 结论

使用简洁明了的语言,描述分析的核心结果。涵盖其能解决的问题,如对说明书撰写所提供的支持、用药方案调整等。

5.7.9 附录

分析数据的格式；基础模型与最终模型的代码、输出结果；分析步骤；以及关键的图表。

5.7.10 其他

在研究报告提交监管部门审评时，还应考量评审关注的焦点，并确保涵盖相关技术指南中推荐的关键报告要素。美国食品药品监督管理局（FDA）的临床药理和定量药理评审人员，常运用基于问题的评审方法，针对下列问题进行评估：

- (1) 目前提供临床药理学的信息在多大程度上构成了有效性的关键性或辅助性证据？
- (2) 申办方推荐的用药方案是否适用于目标适应症的普通患者群体？
- (3) 基于患者内在因素的亚群体是否需要不同的用药方案和药物治疗管理策略？
- (4) 是否存在具有临床意义的食物与药物或药物与药物间的相互作用？若有，则如何解决？

分析报告应包含以下内容的相关证据或解决方案：

- (1) 支持药物的药理活性；
- (2) 识别不同患者亚组，包括具显著治疗效果的亚组、风险与获益平衡的亚组、具有显著毒性的亚组，或治疗效果处于边缘状态的亚组；
- (3) 对于仅进行了一项充分设计且控制良好的临床试验的药物，利用剂量-效应或暴露-效应分析作为支持性证据，以评估药物的风险获益比；
- (4) 证明剂量方案选择的合理性；
- (5) 确定影响药物暴露和药效的内在和外在因素；
- (6) 应用建模和模拟技术，支持给药方案的选择和确定；
- (7) 为特殊人群和特定协变量亚群（如年龄、体重、肾脏/肝脏功能）的剂量选择和调整提供依据。

5.8 质量控制

5.8.1 定义

良好的质量控制（Quality Control, QC）确保了分析的准确性和可靠性，同时保障了数据、代码和结果的规范记录和保存，使整个分析过程具有可追溯性和重现性。

通常，分析过程的 QC 可以分阶段进行，并采用基于风险的方法来确定 QC 的范围。例如，根据分析过程中可能出现错误的概率、发现错误的难易程度以及错误对结果的潜在影响，对风险进行评估和分级。在 QC 过程中识别出的任何错误都应在修正后，进行部分或全面的再次 QC。这一过程应重复执行，直至所有 QC 均通过为止。一般情况下，QC 人员应与分析人员保持独立。

5.8.1 一般流程

首先，对分析数据执行质量控制（QC），以识别并纠正错误，并评估数据的完整性。然后，对分析数据集进行 QC，以确认数据集的格式和数据的准确性。面对大量数据集时，如果时间和人力资源有限，无法进行全面的数据 QC，则应进行风险评估。

在模型开发完成后，需要对控制文件和输出文件进行 QC。在此阶段，应对分析过程或内容进行审查，以确保分析过程遵循了分析计划。在模型开发过程中，可能涉及数百个模型，但由于时间限制，通常不可能对每个模型都进行 QC。因此，通常专注于报告正文中讨论的模型，包括基础模型、全量模型、最终模型和模拟所用的模型等关键模型。

在最终阶段，进行分析报告的 QC 至关重要。此阶段涉及对分析结果图表和数据列表的详尽审查。若分析内容及报告需提交给监管机构，必须确保数据和报告符合相关法规要求。此外，推荐实施同行评审，以进一步评估报告的质量、内容、研究方法、分析的准确性和合理性。同时，建议由多专业背景的人员包括专业的质量控制人员、医学专家和定量药理分析人员，以全面审查报告。

在 QC 过程中，发现的任何错误都应评估其对建模过程和结果的潜在影响。对于不影响建模结果的小错误（例如，药物浓度值的记录错误），建议仅在报告的附录中注明。然而，如果错误对建模结果有重大影响（例如，关键协变量的错误），则必须重新运行部分或全部模型。对于模型结果分析和报告中的错误，必须重新绘制图表，并及时更新报告内容。

1 附 录 A 筛选协变量的方法

2

方法	定义	优点	缺点
全模型估算法（full model estimation, FME)	同时拟合所有协变量，并作出推论	<ul style="list-style-type: none">• 一次拟合略去协变量建模过程• 方便解读输出过程• 一步可得最终模型	<ul style="list-style-type: none">• 必须在事前仔细分辨协变量的效应
逐步法（Stepwise covariate modelling, SCM)	包括正向纳入与逆向剔除二个步骤	<ul style="list-style-type: none">• 方便实施(可使用PsN中自动化处理工具)	<ul style="list-style-type: none">• 选择性偏差• 多次未经矫正的比较• 最终模型之前要得出一个全协变量模型
瓦尔德似然比检验近似法 (Wald’s approximation to the LRT, WAM)	通过全协变量模型获得所有限制模型的LRT 值。最终模型的确立基于 NONMEM 计算的SBC 值	<ul style="list-style-type: none">• 协变量较多时，估算拟合次数较少	<ul style="list-style-type: none">• 需要全协变量模型的协方差矩阵
广义加和模型（generalized additive modelling, GAM)	通过广义的线性加法，将个体参数的贝叶斯预测值与协变量进行回归	<ul style="list-style-type: none">• 概念简单、实施便捷	<ul style="list-style-type: none">• 无法避免收缩和参数间的相关性• 无法处理随时间变化的协变量
拉索法（least absolute shrinkage and selection operator, LASSO)	所有协变量须标准化，使其均值为 0，标准差为 1	<ul style="list-style-type: none">• 在数据集较小时，比SCM 能获得预测能力更佳的模型	<ul style="list-style-type: none">• 处理大数据集时无优势

3

附录 B 质量控制要点

B1 数据

一般错误
<ol style="list-style-type: none"> 1. 拟分析的临床研究是否与纳入群体分析的数据集保持一致？ 2. 所有符合群体分析标准的受试者是否都被纳入分析数据集？ 3. 每个研究对象的给药事件发生时间（日期）是否记录在首个观测值之前？ 4. 除备注行外，单元格是否含有字母或非数字字符？ 5. 关键变量，如自变量和因变量，是否存在空白单元格，未填写任何值？ 6. 是否有不合理的值出现，例如体重和年龄出现负数？ 7. 每个研究对象中不随时间改变的变量（如编号、性别等）是否保持一致？
数据格式和单位
<ol style="list-style-type: none"> 1. 是否合理选取了变量的有效数字？ 2. 每列数据的单位是否保持一致？若数据源自不同的研究或多个研究中心，浓度、剂量以及协变量等数据的格式和单位是否统一？若存在不一致，是否已经恰当地转换为研究计划中指定的单位？ 3. 剂量和浓度的单位是否匹配？如果不匹配，观察室比例因子（Fn）的处理是否恰当？ 4. 当药物以盐的形式应用时，AMT 中的剂量是否已经根据摩尔质量进行了校正？ 5. 在同时对原性药物及其代谢产物进行建模分析时，AMT 值是否已经进行了摩尔质量校正？ 6. 在同时对大分子药物及相关受体建立靶点介导的药物处置（TMDD）模型分析时，AMT 值是否已经进行了摩尔质量校正？ 7. 在合并多个研究的数据时，药物浓度或实验室指标的检测方法是否保持一致？如果不一致，是否已经进行了标准化处理？ 8. 当 RATE 通过 AMT/DUR 求算时，时间单位是否正确。
自变量
每个受试者中自变量（时间）是否按照从先到后的顺序排列？
因变量（DV）
<ol style="list-style-type: none"> 1. DV 值是否为同一类型？若否，则数据集中应有标识 DV 类型的指示变量，并且 NONMEM 中使用此指示变量计算。 2. 若 DV 值包含不同类别的数据，可以通过绘制 DV 值与指示变量的散点图来检查指示变量。如果不同指示变量下的 DV 值出现重叠，则说明指示变量未能正确区分不同类别的 DV 值。 3. DV 值中是否存在离群值(outlier)？如果存在，则应检查数据的准确性。在分析计划中应明确离群值的定义。通过基于分类变量或连续变量的分组，绘制栅栏图，有助于识别离群值。
协变量
<ol style="list-style-type: none"> 1. 每个协变量的范围（包括最大值和最小值）是否都符合生理学标准？ 2. 协变量是否表现出时间依赖性？是否考虑了基线值？所有患者是否遵循了统一的规则？ 3. 随着时间推移，是否观察到协变量发生显著且非连续的波动？这些波动可用于辨识错误或异常值。 4. 不随时间改变的协变量是否会随时间变化？这可以通过按 ID 绘制每个协变量经时变化的栅栏图，并观察是否存在从一个时间点到另一个时间点的“跳跃”。 5. 协变量的直方图分布是否呈现多峰形态？如是则需要核实是否正确，或是否与单位错误标记有关。例如，一些研究对象的身高单位可能是厘米，而其他对象的单位可能是米。 6. 分类变量是否具有适当数量的类别？例如，如果一个变量有四个类别，则在数据集中至少应有四个不同的值。

日期
<ol style="list-style-type: none"> 研究日期是否与试验设计相符？例如，若在 2010 年 1 月 1 日进行给药，而 24 小时后的采样日期标记为 2011 年 1 月 2 日，则存在错误。必须核查时间与日期的一致性。 所有日期和时间的格式是否统一？由于不同国家和地区可能采用不同的日期表示方法，例如 2012 年 4 月 3 日，在欧洲可能被记录为 3-4-2012，而在美国则为 4-3-2012。在合并数据集时需尤为注意。 计算首次给药后时间的方法和单位是否恰当（例如，使用小时或天等）？
缺失数据和衍生数据
<ol style="list-style-type: none"> 是否存在值为零的协变量？这些值是否准确无误？ 缺失值是否已按照分析计划进行合理处理？ 当数据被删除时（如异常值），是否已正确地对其进行注释，并在报告中记录了删除的具体原因？ 对于 BLQ 数据，是否已按分析计划进行处理？例如，设定为零或 LLOQ 的一半等方法进行填补。 衍生变量是否已按照分析计划，采用合理的方法进行计算？
NONMEM 软件的特定用法
<ol style="list-style-type: none"> 是否已在 EVID 列中恰当地标记给药记录（例如，EVID=1）？ 非给药事件是否已在 EVID 列中恰当地标记（例如，EVID=2）？ 是否正确处理了 DV 值缺失的情况？在最简单的情况下，DV 值缺失时，是否已在 MDV 列中用 1 进行标记？ 给药记录是否恰当地反映了研究设计？例如，在使用 ADDL 和 II 时，是否指代准确，并正确使用了时间单位？例如，II 和 TIME 应使用相同的时间单位。 CMT 是否与事件发生的房室相对应？例如，对于一级吸收模型，吸收室 CMT=1，中央室 CMT=2。对于静脉输注给药或零级吸收药物，RATE 是否正确赋值？例如，静脉输注 AMT=1000，RATE=-2，将持续输注时间定义为模型参数并添加变量 DUR 表示，控制文件中设置 D1=DUR。

B2 控制文件、输出文件和数据后处理

Fortran 编译器和第三方软件
<ol style="list-style-type: none"> 1. 分析时使用了何种 Fortran 编译器？ 2. 运行 NONMEM 安装目录下的 CONTROL5, 检查输出结果是否与标准输出结果一致： <ol style="list-style-type: none"> (1) 运行 CONTROL5, 是否出现错误？最小目标函数值是否为 104.561？ (2) THETA(1)、THETA(2)和 THETA(3)的值是否分别为 2.77, 0.0781, 0.0363？ (3) OMEGA(11)、OMEGA(22)和 OMEGA(33)的值是否分别为 5.55, 2.4E-4 和 0.515？ (4) SIGMA(1) 的值是否为 0.388？ 3. 否已正确安装并使用其他第三方处理软件, 例如 PDx-POP、PsN、Pirana 等？
控制文件
<ol style="list-style-type: none"> 1. 是否确保\$DATA 指定文件中的数据列数与输入数据集保持一致, 并且数据列的顺序与\$INPUT 指定的数据项顺序相匹配？ 2. ADVAN 和 TRANS 选项是否与预期的模型和参数设置相符合？例如, ADVAN1 对应静脉给药的一房室模型, 而 ADVAN2 对应口服给药的一房室模型。 3. 比例因子 (scale value) 和采样房室 (sampling compartment) 的容积单位是否保持一致？例如, 是否满足 $S2=V2$ 的关系, 或者是否应用了比例因子以确保剂量和血药浓度值单位的一致性？例如, 如果剂量单位为 mg, 体积单位为 L, 浓度单位为 ng/ml, 则 ADVAN2 的 S2 应表示为 $S2=V2/1000$。 4. 在统计学模型中, 如果 ETA-EPS 存在相互关系, 是否在\$EST 中启用了 INTERACTION 选项？ 5. 在使用微分方程编写控制文件时, CMT 的值是否与房室相对应 (例如, 1 代表给药室, 2 代表口服给药模型的观察室)？ 6. 在\$SUB 语句中的 TOL 值是否大于\$EST 语句中的 NSIG 值？ 7. 如果残差模型采用双向转换法 (TBS, transform-both-side), 是否确保因变量和残差模型使用了相同的转换方法？例如, 如果 TBS 采用对数转换, 则因变量和残差模型是否分别为 $Y=f(\text{LOG}(F))$, $\text{EPS}(n)$ 和 $Y=\text{LOG}(F)+\text{EPS}(1)$。 8. 在使用多个\$EST 语句的情况下, 估算方法及相关选项是否已在报告中明确体现？ 9. 模拟时应关注以下问题： <ol style="list-style-type: none"> (1) MSF 文件是否被用作输入参数？控制文件是否使用了 TRUE=FINAL, 或者是否将参数固定为最终模型的参数, 并且这些参数值是否正确？ (2) 是否已经定义了模拟的次数 (subproblems)？ (3) 是否已经定义了模拟的种子数 (seeds)？ (4) 是否使用了正确的概率密度函数 (probability density function), 如正态分布或均匀分布？ (5) 后处理所需的参数是否已在\$TABLE 语句中完整列出？
输出列表
<ol style="list-style-type: none"> 1. 列表文件中呈现的个体总数和观察事件总数是否与数据集保持一致？ 2. 列表文件内是否含有任何警告或错误信息？ 3. 列表文件所展示的观察事件数量是否与数据集中未被注释的事件数量相符？ 4. \$COV 步骤是否已正确执行？若执行失败, 模型是否仍可视为可接受？ 5. 最终参数估算的有效数字是多少？ 6. THETA、OMEGA 和 SIGMA 的估算值是否在合理范围内？ 7. 是否已经估计了标准误？ 8. 收缩 (shrinkage) 是否在可接受的范围内？ 9. 所有 ETABAR 的值是否在统计学上均无显著意义？ 10. 最终模型的条件数 (conditional number) 是否小于 1000？

11. 列表文件中记录的起始和终止日期时间是否与输出文件中的日期时间相匹配?
输出文件
1. 若未启用 FIRSTONLY 选项, 输出文件中的数据行数是否与录入数据集中的行数相匹配, 同时减去在 \$INPUT 中因 IGNORE 或 ACCEPT 指令而被排除的行数? 2. 输出文件与控制文件中的时间戳是否在时间与日期上保持一致? 3. 输出文件与输入文件的数值精度是否保持一致?
数据后处理
1. 数据是否被后处理软件正确解读? (1) 读取的数据行数是否与输出数据集的行数相匹配? (2) 变量数量是否与输出数据集的列数相符? (3) 原始数据是否与输出文件的数值精度保持一致? (4) 输出数据集中的变量, 如 CWRES , 是否与后处理软件中读取的数据相匹配? (注: 变量名称可能不完全相同, 但数据必须一致) 2. 核查报告中的图: 是否使用了正确的控制文件, 并且使用适当的数据进行绘制? (1) 如果数据后处理软件是基于脚本的, 脚本是否有适当的注释? 包括运行日期、控制文件的名称, 以及用于绘制图表的 NONMEM 输出文件的路径。 (2) 如果数据后处理软件是基于图形用户界面的, 是否有稽查追踪记录, 以确保使用正确的 NONMEM 输出文件来绘制图表? 3. 当 MDV=1 时, DV 、 PRED 和所有残差值是否被设定为 0, 并且这些值在后处理前是否被正确地重新标记为缺失? 4. 是否使用了正确的残差值进行绘图? 例如, 报告中的 CWRES 是指 CWRES 还是 WRES ?

* 以上内容以 **NONMEM** 软件为例, 其他软件可参考。

B3 分析报告

报告结构
<ol style="list-style-type: none"> 1. 报告应该足够详细，确保能被监管者二次评估。 2. 建模过程中的每个假设和决定均应记录、讨论和证明。 3. 报告应该具有高质量，从而“可以认为最终模型很好的描述了数据的特征，因为结果和结论是正确和有效的”。
报告质量
<ol style="list-style-type: none"> 1. 如有可能，重新运行报告中的 NONMEM 控制文件，并比较结果，以验证报告中的结果与 NONMEM 输出结果是否一致。 2. 报告是否包含了 NONMEM 和 Fortran 编译器的版本信息？ 3. 在数据分析过程中使用的第三方软件，例如 PDx-Pop 和 PsN 等，是否标明了软件的版本？ 4. 报告中是否涉及数据分析计划之外识别出的任何异常值？ 5. NONMEM 中选择的 ADVAN 与报告中的模型是否匹配？例如，ADVAN 是否对应一室模型，而 ADVAN2 是否对应有口服吸收的一室模型？ 6. 协变量模型的函数形式是否与报告中描述的一致？即控制文件中的协变量模型与报告中的模型是否相同。 7. 控制文件中指定的估算方法是否与报告中使用的估算方法一致？ 8. 最小化过程是否成功完成？报告中是否记录了任何错误和警告？ 9. 报告是否包含了收缩（shrinkage）？如果收缩超过了预定义的临界值，是否在报告中进行了说明？ 10. 报告是否包含了 ETABAR 统计结果？是否记录了 p 值小于 0.05 的情况？ 11. 报告是否提供了所有可估计模型参数值及其标准误？ 12. 报告中是否以统一的形式报告了方差项（variance），例如方差的百分比系数或方差的对数值？ 13. 最终模型是否报告了条件数？ 14. 报告附录中的所有模型控制文件和输出文件是否与报告正文保持一致？ 15. 诊断图是否与模型相匹配？ 16. 如果进行了模拟，报告中是否记录了种子数和模拟次数？ 17. 所有包含模型运行概况的表格是否与 NONMEM 列表文件中的信息保持一致？
报告内容
<p>早期</p> <ol style="list-style-type: none"> (1) 是否拥有充分的内容以支持和解决问题？ (2) 是否包含监管部门所需的信息？ (3) 文档是否具备说服力和逻辑性？ (4) 否具备充分的研究背景以阐明研究主题？ (5) 关键信息是否在突出的位置呈现？ (6) 文档的设计和视觉效果是否令人满意？
<p>晚期</p> <ol style="list-style-type: none"> (1) 数据是否完整准确、且保持一致性？ (2) 所有附图是否均已明确标注并易于理解？ (3) 文档格式是否保持统一？ (4) 语言表达是否清晰且正确无误？ (5) 是否已经妥善解决所有的分歧和矛盾？
其他

- (1) 部分文档已被审核（随后进行修改与批准），则在之后的审阅中，应避免被重复审阅。
- (2) 为确保文档审核的效率和质量，已审核的文档应避免在后续审阅中被重复检查。
- (3) 应避免总是从文档的开头部分开始审阅，可以从方法论或结果部分入手，防止对文档开头部分的过度关注。
- (4) 报告的撰写应在分析结束之前，甚至可以实时进行。研究目的、研究方法、人口统计学特征、基本方法等部分可以预先准备，同时为已确定的图表在文档中预留空间。分析完成后，只需将结果、解释和结论插入到预先准备好的文档中。
- (5) 在报告的某些部分完成时，即可进行阶段性审阅，而不是等到整个报告完成后才开始审阅。如果无法实时撰写报告，建议与团队成员讨论结果，并就结果的解读、结论和临床意义等达成共识。
- (6) 建议向审阅者提供具体的意见，而不仅仅是截止日期。例如，可以指出审阅者无需专注于方法学，而应重点关注方法流程、结果和讨论部分。同时，可以明确告知哪些部分尚处于草稿阶段，不适宜审阅，或者哪些部分不需要审阅。
- (7) 培训审阅者如何进行复审并提出建设性建议，如何从作者的角度审阅并提供修改意见，避免模糊的建议，如“请修订”或“？”。审阅者应详细阐述其建议，并提供明确的细节。
- (8) 建议制定评阅检查清单，对文档整体提出具体意见。应避免将重点放在语法和拼写上，而应提出整体建议和评论，以提升报告的质量。例如，可以指出“模型显示年龄影响清除率，但缺乏图表支持此结果”。
- (9) 建议采用文档共享编辑审阅系统，以管理审阅、修改意见和文档版本，从而提高工作效率。

参 考 文 献

- (1) 国家药品监督管理局.《模型引导的药物研发技术指导原则》，2020年12月.
- (2) 国家药品监督管理局.《群体药代动力学研究技术指导原则》，2020年12月.
- (3) European Medicines Agency, EMA. Guideline on reporting the results of population pharmacokinetic analyses. 2007年6月
- (4) U.S. Food and Drug Administration, FDA: Population Pharmacokinetics Guidance for Industry. 2022年2月
- (5) 日本（PMDA）： Guideline on Population Pharmacokinetic and Pharmacodynamic Analysis. 2022年2月
- (6) Guidelines for the quality control of population pharmacokinetic-pharmacodynamic analyses: an industry perspective. AAPS J. 2012;14(4):749-58.
- (7) Establishing best practices and guidance in population modeling. CPT Pharmacometrics Syst Pharmacol. 2013;2(7):e51.
- (8) 新药研发中群体药动学/药效学研究的一般考虑.中国临床药理学与治疗学, 2019, 24(11):1201-1220.