



国际人用药品注册技术协调会

ICH协调指导原则

儿科外推

E11A

终版

2024年07月 31日采纳

该指导原则由相应的 ICH 专家工作组制定，按照 ICH 进程，已通过药品监管机构讨论。在 ICH 进程的第 4 阶段，最终的指导原则草案被推荐给 ICH 区域的监管机构采纳。

E11A

文件历史

编码	历史	日期
E11A	在第2阶段由ICH大会成员认可，公开征询意见。	2022年4月4日
E11A	由ICH大会监管成员在第4阶段批准。	2024年8月

法律声明： 本文件受版权保护，在始终承认 ICH 版权的前提下，除 ICH 标识外，基于公共许可，可以被使用、复制、在其他作品中引用、改编、修改、翻译或传播。如对本文件进行任何改编、修改或翻译，必须采取合理措施来清晰地标明、界定或以其他方式标记对原始文件所作的更改。任何暗示 ICH 授权或支持对原始文件的改编、修改或翻译行为必须避免。

本文件根据现有内容提供，不作任何保证。任何情况下，ICH 或原始文件作者不会对任何由使用本文件造成的索赔、伤害或其他责任负责。

上述许可不适用于第三方提供的内容。因此，对第三方拥有版权的文件需获得版权所有人的复制许可。

ICH 协调指导原则

儿科外推

E11A

ICH共识指导原则

1. 引言	4
1.1 本指导原则的目的	4
1.2 背景	4
1.3 范围	5
1.4 一般考虑	6
2. 儿科外推框架	8
3. 儿科外推概念	10
3.1 疾病	10
3.2 药物药理学	13
3.3 治疗反应	14
3.4 安全性考虑	16
3.4.1 安全性外推	16
3.4.2 额外安全性考虑	18
3.5 现有数据的来源和类型	19
3.6 整合证据并确立儿科外推概念	22
4. 儿科外推计划	25
4.1 一般考虑	25
4.1.1 将青少年纳入成人试验	26
4.1.2 建模和模拟方法	27
4.1.3 剂量选择	30
4.1.4 剂量范围数据的使用	32
4.1.5 生物标志物的使用	32
4.1.6 建立参照人群和目标人群之间不同终点的关系	33
4.1.7 安全性外推计划	34
4.2 儿科外推计划研究设计方法	35
4.2.1 暴露量匹配方法	36
4.2.2 PK/PD方法	39
4.2.3 有效性研究	40
4.2.3.1 单臂有效性研究	42
4.2.3.2 外部对照研究	42
4.2.3.3 同期对照有效性研究	43
4.2.3.4 纳入外部数据	44
4.2.3.5 量化使用参照数据的影响	45
4.2.4 儿科试验的呈现和合理性依据	47

4.2.5 分析、报告和解释.....	48
4.2.6 在儿科试验分析中利用参照数据的方法.....	50

1. 引言

1.1 本指导原则的目的

本指导原则旨在为使用儿科外推提供建议并促进国际协调，以支持儿科药物的研发和许可。儿科外推法的协调应降低不同地区出现明显差异的可能性。重要的是，协调还应减少儿科人群暴露于不必要的临床试验，并促进其在全球范围内更及时地获得儿科药物。使用“应”一词意味着建议或推荐某事，但除非全球监管机构有具体的法律或法规要求，否则并非必须遵守。

1.2 背景

不同的监管机构先前已发布了讨论儿科外推的区域性指导原则。ICHE11(R1) 指导原则将儿科外推定义为：“导一种为儿科人群安全和有效使用药物提供支持证据的方法，其前提是假设儿科人群[目标]和参照人群（成人或其他儿科人群）的疾病进程和药品的预期治疗反应足够相似。”参照人群可以包括其他儿科年龄亚群。儿科外推可以根据对两个人群的疾病¹、药物药理学和治疗反应的相关相似性评估，将关于参照人群的已知情况（如药代动力学（PK）/给药方案、有效性和安全性）扩展至目标人群。

¹就本文件而言，“疾病”包括“疾病”和“（异常身体）情况”。（异常身体）情况可能包括患有某种疾病以及处于患有某种疾病的风险之中。

长期以来，安全性外推通常被视为不可接受。然而，我们对参照人群和目标人群在安全性方面的相似性和差异性的理解已经发生了变化。如ICH E11(R1)指导原则所述，使用参照人群中生成的数据来定义应在目标人群中收集数据的范围和程度的原则，也适用于生成安全性数据(见第3.4节)。

本指导原则旨在补充和扩展ICH E11(R1)，为使用儿科外推以优化儿科药物研发提供更全面的框架。ICH E11(2000年版)和E11(R1)(2017年版)应被视为配套指导原则。它们提供了安全、高效和合乎伦理的儿科药物研发的方法指导，包括：(1)何时启动儿科项目；(2)研究类型；(3)年龄分类；(4)年龄适宜剂型；以及(5)伦理原则。ICH E11(R1)还鼓励申办方在研发过程的早期与监管机构进行沟通，讨论优化儿科药物研发的方法，包括使用儿科外推和建模与模拟(M&S)。ICH E11A旨在帮助申办方和监管机构确定儿科外推的应用程度，以及为解决剩余的不确定性和知识缺口，支持儿科人群安全和有效地使用药物而应收集的信息。

1.3 范围

本指导原则为使用外推法作为支持儿科药物研发的工具提供了一个框架。该框架描述了一个有助于了解现有可用信息，确定指导研发所需信息的缺口，和生成额外信息的方法(在需要时)的迭代过程。本指导原则推荐了用以评估影响参照人群和儿科目标人群之间疾病、药物药

理学和治疗反应相似性确定的因素的评估方法。此外，还讨论了疾病特征、药物药理学和治疗反应如何影响相似性确定。

本指导原则讨论了如何利用定量工具（如M&S以及其他统计学方法）来填补知识缺口和/或减少不确定性。本指导原则并非旨在全面列出在儿科药物研发中外推发挥重要作用的所有疾病和/或情况，而是解释了如何实际应用儿科外推来支持产品在儿科人群中的安全性和有效性。本指导原则不会讨论其他类型的“外推”，例如，将一个区域的临床数据外推至另一个区域的人群，以作为药物注册的基础（见ICH E5）。尽管本指导原则中提到或解释了一些定量策略，但这并不意味着会对其进行全面描述。本指导原则的读者应对临床试验研发中使用的定量方法的作用有一些基本理解。

1.4 一般考虑

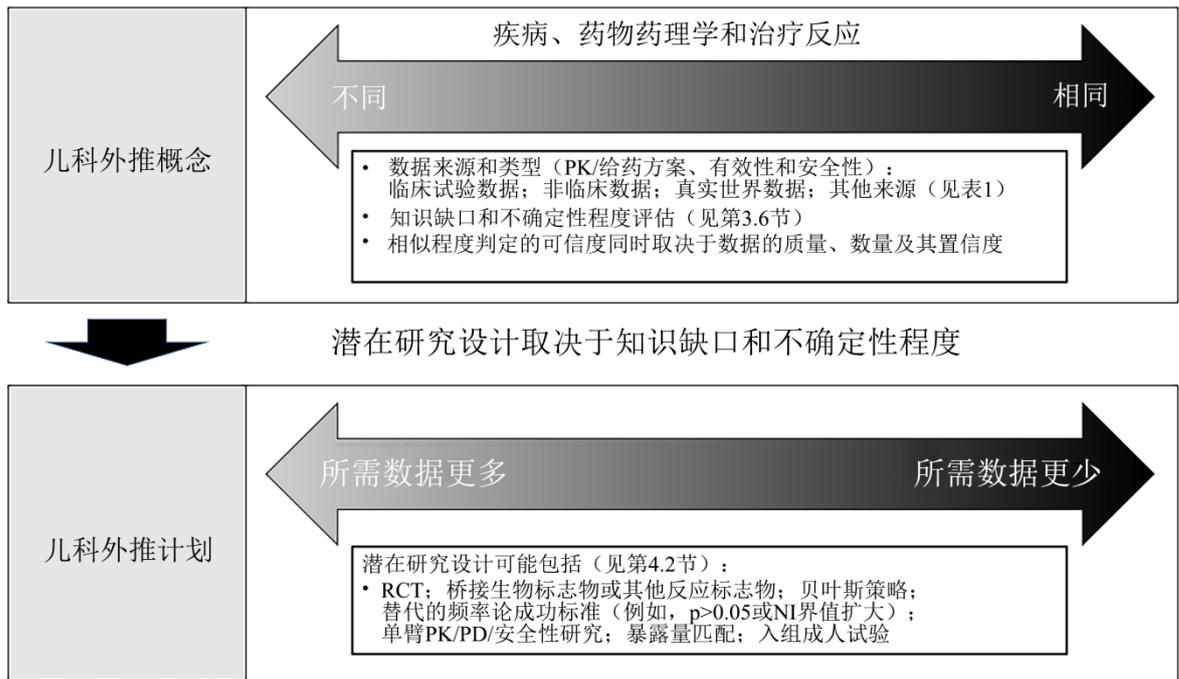
使用儿科外推可确保儿科人群仅在必要时参与临床试验，以增进在儿科使用中对药物的科学认识。根据ICH E11(R1)，需要有充分的预期临床获益以证明让儿科人群暴露于研究药物的风险的合理性。当儿科研究作为成人驱动药物研发的一部分时（包括当监管机构要求开展此类研究时），开展研究的理由通常会隐含地假设参照人群和目标人群之间的疾病存在一定程度的相似性。因此，与成人疾病相关的儿科项目可能适合纳入一定程度的儿科外推。尽管因为快速的生理变化和器官成熟，外推至年幼的儿科人群，特别是新

生儿可能具有挑战性，但本儿科外推框架中的一般原则仍然适用。

在ICHE11(R1)对儿科外推的定义中，“足够相似”可能意味着儿科外推必须超过某一阈值才能被监管机构接受。然而，目标人群和参照人群之间的疾病和预期治疗反应是否可以被认为是足够相似，并不是一个简单的“是或否”问题。因此，本指导原则并未使用离散类别（如完全、部分或无）来描述儿科外推的不同方法，而是倾向于确定研究设计，此类设计能够根据对现有数据的评估，来解决剩余的知识缺口和不确定性。如本指导原则中所述，外推的使用反映了参照人群和目标人群之间可能存在疾病、药物药理学和治疗反应的相似/差异的演变连续性（见图1）。相似性程度的结论得出，在一定程度上取决于对证据强度的多学科评估、所审查数据的置信度以及剩余的知识缺口。重要的是，如图1所示，确定参照人群和目标人群的相似程度不是为了展示一个离散的“相似度量表”，而是提供从概念上理解此框架的指导。一旦审查了参照人群和目标人群之间的相似性，并确定了剩余的知识缺口和不确定性，就可以由多学科团队创建一个儿科外推计划。研究设计、方法和分析的选择取决于知识缺口和需要解决的不确定性水平。外推计划应解决这些不确定性，利用临床判断来确定可接受不确定性的容许水平。儿科外推计划中提出的数据类型和研究设计可以涵盖从暴露量匹配到随机对照

试验（RCT）等多种选择（见图1和第4.2节）。在有限的情况下（例如，已有稳健的临床信息），可能无需在目标人群中收集额外的药代动力/给药方案、有效性和/或安全性数据。

图1：儿科外推的演变连续性



2. 儿科外推框架

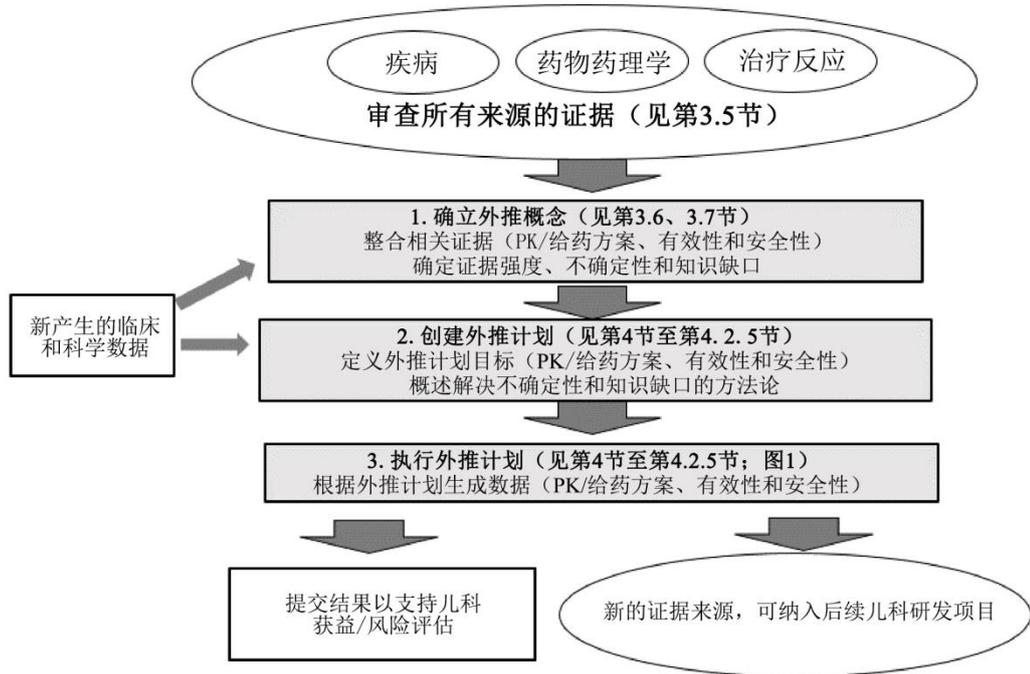
本指导原则进一步完善了最初在ICH E11(R1)中引入的外推框架。该框架由三部分组成：确立儿科外推概念、创建儿科外推计划，以及执行儿科外推计划（见图2）。

第一步是确立儿科外推概念，作为儿科外推计划合理性依据。儿科外推概念评估了参照人群和目标人群之间疾病、药物药理学和治疗反应的相似性和差异性，包括已知和未知的情况。这一概念是通过全面详细审查有关定义疾病的各种因素、药物药理学以及参照人群和目标人群临床治疗反应的

现有信息而形成的。应确定和评估影响参照人群和目标人群治疗效应的因素。对现有知识进行审查后，应综合数据以确立儿科外推概念。数据审查和综合方法可以包括定量方法，如M&S和统计方法（见第3.6节）。数据综合旨在了解证据强度，确定重要的知识缺口，从而得知可能还需要哪些额外数据。

确立儿科外推概念后，应创建儿科外推计划。该计划应包括数据生成目标和方法，以确认做出的假设、解决不确定性和知识缺口，并支持目标人群中的获益风险评估，供监管决策使用。此外，受新出现的临床和科学数据影响，儿科外推概念可能会有所变化。在这种情况下，与其放弃基于先前儿科外推概念的现有儿科外推计划，不如修改儿科外推计划以适应新的科学和临床认知。

图2：儿科外推框架



执行计划还应包括对生成的数据进行评估，以确认做出的假设，并解决通过儿科外推概念识别出的不确定性。

应提交结果以支持儿科风险-获益评估。这些结果应被视为新的证据来源，用于后续儿科研发项目。

3. 儿科外推概念

确立儿科外推概念需要了解在参照人群和目标人群中可能影响疾病、药物药理学和治疗反应以及安全性的因素。

3.1 疾病

评估参照人群和目标人群之间疾病相似性和差异性确实是确立儿科外推概念的一个关键因素。尽管长期以来，儿科外推通常基于对疾病相似性的二元判断（即是或否），但如今对参照人群和目标人群之间疾病相似性和差异性的理解已

经变得更加细致入微（见第1.4节）。对疾病相似性的评估并非旨在确定参照人群和目标人群的疾病是否“完全相同”，而是确定疾病的相似程度和/或差异程度。即便疾病存在差异，仍然可能因为存在一些相似之处而允许使用儿科外推。

即使总体人群中的疾病不够相似，也可能在参照人群和目标人群中找出足够相似的疾病亚组，以支持使用儿科外推。例如，导致成人心力衰竭的诸多病因与儿科的心力衰竭病因并不相似；然而，由扩张型心肌病引起的心力衰竭在成人和儿科人群中是相似的，因此允许从成人向患有扩张型心肌病的儿科患者进行外推。

为增强对不同人群间疾病相似性的理解的信心，评估疾病相似性时还应尝试确定已审查证据中存在的知识缺口和不确定性，并确定需要哪些额外的证据。重要的是，评估疾病相似性不是一个静态或“一次性”的工作。在获得额外的知识后，应将其纳入到儿科外推概念中的疾病相似性评估中（见图2）。此时的评估可能会确认或改变先前的假设，从而对外推计划产生不同程度的影响（无影响或有一定影响）。

评估疾病相似性时需要考虑的因素

在评估参照人群和目标人群之间的疾病相似性时，应审查以下因素：

疾病的病理生理学

应对参照人群和目标人群之间的疾病的病理生理学和

病因进行评估。收集的相关信息可能包括生化、遗传/表观遗传学、细胞学、组织、器官系统和流行病学信息，这些信息描述了参照人群和目标人群之间的相似性和差异性。评估可能还包括确定疾病临床表现的差异是否取决于发病年龄、年龄相关的表型表达或其他年龄相关差异。评估疾病病理生理学中常见的生物标志物（如有），包括疾病进展，通常有助于确定参照人群和目标人群之间的疾病相似性。还应尽可能评估未经治疗疾病的结局的相似性。

疾病定义

应对参照人群和目标人群之间的疾病定义和诊断标准进行评估。在评估参照人群和目标人群之间的相似性和差异性时，应考虑以下问题：

- 定义疾病的临床表现或诊断标准是什么？
- 参照人群和目标人群之间的疾病临床表现的相似程度如何？
- 如何衡量疾病临床表现？
- 是否有相似的衡量方法可用于定义参照和目标人群中的疾病临床表现？
- 参照人群或目标人群中是否存在疾病亚型（如基于严重程度、遗传学、分子标志物等划分的亚型）？
- 参照人群和目标人群中的疾病亚型有何异同？
- 是否需要考虑其他用于定义疾病的因素（如预后因素、

预测因素、遗传/表观遗传学因素、心理社会因素等)？

疾病进程

应对参照人群和目标人群之间疾病进程的相似性和差异性进行评估。评估时应考虑以下问题：

- 参照人群和目标人群中的疾病临床病程有何异同？疾病进程是否会因发病年龄等因素而有所不同？
- 参照人群和目标人群中是否有相似的终点和/或生物标志物可用于衡量疾病进展？
- 参照人群和目标人群疾病的短期或长期结局是否相似，是否可以使用相似方法对这些结局进行衡量？
- 参照人群和目标人群均使用的现有治疗方法有哪些？
- 这些治疗方法对参照人群和目标人群的疾病进程有何影响（例如，相对于疾病发病时间和患者年龄的治疗时机、治疗频率、治疗时长）？

尽管参照人群和目标人群之间疾病进展的频率、严重程度或时机可能不同，但因为疾病进程中存在的某些共性，仍可能允许儿科外推的使用。如果某种治疗方法在参照人群中改变了疾病进程，但尚未在目标人群中获批，这不一定意味着，就儿科外推的目的而言，两个人群之间的疾病进程在当前阶段有所不同。

3.2 药物药理学

用于儿科外推目的的药物药理学评估内容包括：研究药

物的吸收、分布、代谢和排泄（ADME）特性、药效学（PD）（见第3.3节）和研究药物的作用机制（MOA）。

应考虑内在和外在因素对ADME的潜在影响，如体重、体表面积、年龄、器官成熟度、合并用药以及其他相关因素（例如，蛋白结合、代谢酶、转运体、肾功能或剂型选择）。ADME特性的差异可能导致药代动力学（PK）参数（例如，清除率、分布容积）和药物暴露量出现差异。暴露量是一个广义概念，涵盖了在单个时间点测量的药物（原形药物和/或代谢物）全身（或其他生物体内特定组织）暴露量（例如，最大浓度或谷浓度）和/或在一段时间内的暴露量（例如，AUC_{0-t}或平均浓度）。

在评估药物的PD和MOA时，应考虑成熟度相关差异的潜在影响，例如，药物靶点的表达水平和敏感性以及潜在的下效应器（如适用）。这些差异可能导致参照人群和目标人群之间有效性和安全性的暴露量-效应（E-R）关系不同。此外，药物的次要PD特性（即，脱靶效应）的差异也可能导致参照人群和目标人群中不同的毒性特征。

3.3 治疗反应

为了评估治疗反应的相似性和差异性，应对参照人群和目标人群中可获得的知识进行彻底审查，包括治疗相同疾病时对研究药物、其他同类药物以及其他类别药物的反应。同样地，该药物或同类药物在其他疾病中产生的数据也可以作

为评估治疗反应的相似性或差异性的相关知识来源。在该项评估中应包括对参照人群和目标人群的剂量/暴露量和治疗反应的数据的评估（见第4.1.3节）。

评价治疗反应的相似性时应考虑的因素

参照人群和目标人群之间治疗反应的相似程度也可以支持疾病的相似程度。在参照人群和目标人群中与疾病内在相关的靶点（如蛋白质、受体、mRNA等）如果显示出相似的治疗效果，则可以支持疾病的相似性。

评估参照人群和目标人群之间治疗反应的相似性，应包括对剂量/暴露量和治疗反应相关数据的审查。应将发育和成熟度变化对剂量/暴露量和临床反应的潜在影响作为此评估的一部分。应评估对药物靶点的认知以及其在正常发育、疾病病理学和预期治疗反应中的作用。例如，如果某个受体在出生后6个月内不存在，则预期仅靶向该受体的药物在该年龄组不会产生治疗反应。应评估可能在参照人群和目标人群之间存在差异且影响治疗反应的因素（例如，前序治疗、合并用药、伴发疾病、器官功能、遗传构成等），以评估儿科外推的适用程度是否受到影响。此外，用于衡量反应的终点相似性和差异性的认知，可能会影响对治疗反应相似性的总体评估。

在评估治疗反应的相似性时，应考虑以下问题：

- 参照人群和目标人群中是否有相似的终点衡量方式

（例如，临床终点、生物标志物终点、复合终点等）？

- 如果参照人群和目标人群中的反应终点或终点衡量方式不同，那么终点之间的关系是什么（例如，参照人群中的临床终点与目标人群中的生物标志物终点的关系）？

- 是否有同时影响暴露量和反应的因素（例如，基线疾病严重程度、前序治疗）？

当评价治疗反应的相似性时，应考虑是否存在可能导致目标人群和参照人群之间所测反应存在差异的年龄/成熟度相关因素（见第3.1和3.2节）。对于许多儿科药物研发项目，目标儿科人群的主要终点与参照人群的主要终点不同。在这种情况下，可以比较主要终点的一个或多个组成部分和/或次要/探索性终点，以了解不同终点之间的关系。例如，如果有一个与参照人群中已建立的临床有效性终点相关的生物标志物，并且该生物标志物也与目标儿科人群中的临床有效性相关，则这样的“桥接生物标志物”可以支持治疗反应的相似性（见第4.1.5和4.1.6节）。

3.4 安全性考虑

在其他ICH指导原则（如E2、E6、E9、E11、E11(R1)）中对制定总体安全性数据收集和不良事件报告计划的基本考虑进行了讨论。本节描述了与安全性外推有关的特定考虑，作为确立整体儿科外推概念的一部分。

3.4.1 安全性外推

适当使用参照人群中生成的数据来定义需要在目标人群中收集的有效性数据范围及程度的基本原则也适用于安全性数据的生成（见第4.1.7节）。可以考虑根据参照人群中与目标儿科人群相关的已知和/或潜在安全性问题的现有知识，对安全性数据进行外推。应考虑将其他相关的信息来源作为本分析的一部分（见第3.5节）。这些数据应有助于提高药物在特定儿科人群中预期安全性特征的确定性，并确定是否需要在儿科项目中解决额外的知识缺口。在外推概念和计划中，应包括对安全性的外推适用性和程度的评估。

参照人群和目标人群之间安全性特征的相似性和差异性应被视为具有演变连续性。应在药物研发规划的早期考虑用于支持将安全性数据外推到目标人群的安全性数据的来源及数量。参照人群可能包括暴露于同一药物或同类药物的儿科和/或成人人群。数据也可来自于使用了不同给药方案和/或针对不同疾病/适应症接受了治疗的参照人群。例如，ICH E11和E11(R1)中定义的青少年安全性数据收集，也可作为在更年幼患者中支持安全使用药物的新证据来源。青少年入组成人试验，或与成人试验同时进行，可能有助于更早评估青少年群体的安全性（见第4.1.1节）。

在确立安全性外推概念时，应考虑以下问题：

- 作为安全性外推的一部分，拟研究的目标儿科人群的年龄范围是多少？

- 从参照人群中获得的可用安全性数据的数量/质量如何？
- 研究药物是否存在与儿科安全性相关的已知靶向或脱靶效应？

- 是否需要数据来说明目标儿科人群中年龄特异性的短期和长期不良反应（这些不良反应可能尚未在参照人群的研究中识别）？

- 与目标儿科人群相比，参照人群的预期治疗持续时间和治疗效应大小如何？

- 参照人群和目标儿科人群中的预期药物暴露量是否可比？是否能够通过达到特定PD效应或临床反应所需的暴露量来预测目标儿科人群中的特定毒性？

- 从非临床来源（见表1）获得的哪些已知信息可用于目标人群？

- 参照人群和目标人群之间是否存在其他可能限制安全性外推的差异（例如：未在参照人群中使用，但在目标人群中使用的背景治疗可能会增强安全性信号；参照人群使用制剂中的辅料）？

3.4.2 额外安全性考虑

作为确立外推概念的一部分，在对安全性外推进行评估后，可能需要在已经收集的数据基础上收集额外的安全性数据。当目标人群中存在未解决的知识缺口和/或特定年龄相关的安全性问题（例如，皮质类固醇对有开放骺板的青春期前

儿童存在生长速度降低的影响)时,可能需要收集额外的安全性数据。因此,在获得批准后,仍可能需要在儿科目标人群中进行更长期的安全性数据收集。

在某些情况下,作为确立外推概念的一部分,应考虑额外收集儿科安全性数据。例如:

- 当药物是新分子实体,属于新药类别时
- 当存在已知的与年龄相关的靶向或脱靶安全性问题时
- 当参照人群中存在重要的安全性发现,对于儿科临床具有特殊重要性时
- 当药物具有窄治疗指数时

最终,应收集的安全性数据的类型、数量和时机将取决于在目标人群中确认的安全性相关的知识缺口(作为外推概念的一部分)。此外,基于相关现有数据的临床论证应作为确定安全性数据集大小的依据;不鼓励随意设定安全性数据集的大小。建议尽早与监管机构进行讨论。

3.5 现有数据的来源和类型

现有数据的使用应符合目的(即,生成数据的场景适用于预期使用数据的场景)。评估参照人群和目标人群之间的相似性和差异性时,数据的数量和质量是至关重要的考虑因素。应使用所有相关数据来确立外推概念并创建外推计划。此类信息还可能包括来自正在进行的成人/儿科研发项目的数据,或来自终止项目的相关数据。表1中包括了数据来源和

类型的示例，并将在本节中作进一步讨论。此表并非旨在提供一个全面的列表，因为还可能涉及其他来源和类型的数据。

表1.评估疾病、药物药理学和治疗反应的数据来源和类型示例

数据来源	数据类型
临床试验数据	药物或同类药物在相同疾病中的PK、PD、E-R和临床数据
	药物或同类药物在其他相关疾病中的PK、PD、E-R和临床数据
	不同类药物在相同疾病中的PK、PD、E-R和临床数据
非临床试验数据	获得自动物模型的ADME数据
	计算机模拟、体外和体内动物数据（如动物疾病模型、PK、PK/PD、MOA）
	成年和幼龄动物毒理学数据
真实世界数据	包括但不限于疾病登记（地区、国家、国际）、电子健康记录、健康索赔数据库
其他来源	包括但不限于系统综述或荟萃分析，相关发表文献
	专业组织指南/临床实践指南/共识文件
	已发表的模型/模拟（如PK/PD、机制性）
	专家意见
	标准治疗/医学实践

临床试验数据

应对患相同疾病或相关疾病的人群的临床试验数据（例如：获自对照试验、前瞻性观察性研究、PK、PK/PD和/或生物标志物研究）进行评估，以了解参照人群和目标人群之间的相似性和差异性。应评估所有相关的临床试验数据，包括进行中和已完成的研究、发表和未发表的研究（无论结果是阳性还是阴性）。

非临床试验数据

还应对非临床来源（如体内、体外和计算机模拟模型）的数据进行评估（如可用）。这些数据可能包括PK、PD和/或疾病模型。一般来说，当临床数据可用时，获自动物模型的数据可能就会不太相关，但情况并非总是如此。在某些情况下，只有非临床数据能支持疾病相似性，尤其是无法收集临床数据时（例如，炭疽或鼠疫）。

真实世界数据（RWD）

RWD在支持儿科外推（包括儿科外推概念和计划）方面的可用性正在不断提升。因此，应与监管机构适当讨论RWD用于支持儿科外推的充分性、相关性和程度。在确立儿科外推概念时，应考虑对RWD来源的数据（包括但不限于电子健康记录、索赔数据库和登记记录）进行审查。

其他来源

专家意见（包括专业组织制定的临床实践指南）可用于支持外推概念。专业组织发布的临床实践指南被认为比未发表的专家意见更具信息价值。然而，由于标准治疗存在差异，不同地区发布的指南和专家意见可能有所不同。在未评估证据强度的情况下，依赖专家意见或标准治疗通常是不够的（见第3.6节）。

总之，上述数据的来源和类型各有优缺点。要确定支持参照人群和目标人群之间相似性的数据来源和类型的置信

度，需要对每个来源的数据数量和质量以及评价数据的背景进行评估。应对所有相关数据进行审慎的和多学科的评估，以证明使用证据支持外推概念的合理性。

3.6 整合证据并确立儿科外推概念

确立儿科外推概念的目标不仅仅是确定儿科外推的可接受性，还需要描述所做的假设，详细说明所有知识缺口，并评估可用证据中不确定性的影响。本节为儿科外推概念中应包含的信息审查、综合和呈现提供了指导。

整合现有证据

整合现有证据涉及全面审查，以评价参照人群和目标人群之间疾病、药物药理学和治疗反应的相似性和差异性（见图2）。一旦对证据进行了审查和整合，则应对证据强度进行评价，并确定证据中的缺口。整合证据应解决以下问题：

- 证据体系是怎样的？证据的临床相关性是怎样的？
- 证据的强度和局限性有哪些？
- 不同来源和类型的数据的结果一致性如何？
- 数据中存在哪些不一致性，这些不一致性如何影响相似性评估？

这些问题的答案将提示在最终确立外推概念之前需要哪些额外的信息（如有）和/或外推计划中应收集哪些额外的数据。

可用于整合证据的方法

应使用定量综合来整合现有数据(见表1)以确立外推概念。为了充分整合多样的数据类型和来源,可以采用多种定量方法(经验性和机制性的方法),并应进行记录。鉴于整合的关键预测因子(内在/外在因素)以及变异的来源和类型(例如,个体间、个体内和场景间)可能存在潜在差异,可能有必要采用多种方法来准确利用所有数据。系统生物学/药理学模型可用于评估和预测疾病生物学、病理生理学和治疗反应。群体建模(见4.1.2节)可用于指导对关键参数的评估,并量化变异性的来源。荟萃分析技术可用于综合来自多个来源的有效性和安全性数据。

有多种方法可用于定量评估不同人群中疾病和/或治疗反应的相似性(见4.1.2节)。应依据相似性评估数据来选择合适的方法。例如,当使用频率论方法时,可以通过比较点估计值及其相关置信区间来指导参照人群和目标人群之间的反应相似性评估。在多数情况下,不适宜仅通过置信区间的重叠来确定相似程度。评估还应考虑点估计值的准确度及其之间差异的大小。贝叶斯层次模型也可以用于整合和综合可用证据。对于定义、规定和在评估中如何对不确定性加以考虑的方式,以及用于评估疾病和/或反应相似性的任何模拟,都应该进行记录。此外,应具体说明与不确定性的定义或表达有关的任何相关假设。

也可考虑对现有数据进行其他探索性分析,以评估相似

性。例如，如果在参照人群中开展的试验纳入了各个年龄组，可以考虑评估每个年龄组的反应一致性。用于评估亚组间反应一致性的方法，在其他ICH指导原则中已有描述（见ICH E17）。

在评价参照人群和目标人群之间的疾病和/或反应相似性时，鉴于数据中固有的不确定性，现有数据可能无法得出明确的结论。因此，建议申办方酌情与监管机构讨论拟定方法的可接受性。

确定知识缺口

整合现有证据后，应随即确定知识缺口。在创建儿科外推计划前，可能需要基于新产生的临床和科学数据，解决其中的某些知识缺口（见图2）。然而，某些知识缺口并不一定会妨碍儿科外推计划的建立。儿科外推计划中应说明何时需要收集哪些数据以填补这些知识缺口。确定知识缺口时应解决以下问题：

- 已确定的知识缺口有哪些？
- 在创建儿科外推计划之前，是否需要为这些知识缺口收集额外的数据？如果是，何时以及如何收集这些数据？
- 如果这些知识缺口不妨碍儿科外推计划的创建，那么应何时以及如何儿科外推计划中解决这些缺口？

3.7 儿科外推概念的确立

确立儿科外推概念时应包括对参照人群和目标人群之

间的总体相似性和差异性、当前知识缺口、不确定性和数据局限性的总结。应包括以下内容：

- 对参照人群和目标人群(疾病、药物药理学、治疗反应)之间相似性和差异性证据的评估(即总体优势和劣势)。还应包括对证据数量和质量的评价。

- 评估现有安全性信息,以及该安全性信息如何影响外推概念。

- 评估知识缺口及其如何影响外推概念的置信度和不确定性。此外,该评估应描述何时以及将如何解决知识缺口。

4. 儿科外推计划

确立了儿科外推概念后,应在外推计划中详细说明拟开展的研究和/或数据分析(包括基于模型的分析),以及开展它们的理由依据。研究的设计和/或分析应反映出参照人群和目标人群之间已经确定的相似性和差异性,以及为解决外推概念中的知识缺口所需要收集的必要信息。

4.1 一般考虑

作为创建儿科计划的一部分,有一些一般考虑,可能适用于任何选择的研究设计(见第 4.2 节)。一些重要的一般考虑包括:将青少年纳入成人试验;使用模型引导方法;剂量选择的考虑因素;使用剂量范围数据;使用生物标志物;在参照人群和目标人群之间终点不同时关于终点选择的考虑,以及安全性计划的创建。本节将对上述考虑因素进行讨

论。在创建儿科外推计划时，应尽早考虑这些问题。

4.1.1 将青少年纳入成人试验

将青少年纳入成人临床试验可以使青少年更快获得安全有效的治疗，并加速所需儿科数据的收集。长期以来，儿科试验通常在面向成人的研发完成后和/或药物已获批用于成人后才开始进行。药物在儿科为超说明书使用，因此儿科试验入组可能会很慢，进一步延迟了更多儿童和青少年获得有效治疗的时机。将青少年纳入某些适宜疾病和/或靶点的成人试验可能会解决这一问题。如果将青少年中的研究结果，用于桥接成人有效性和/或安全性外推至其他儿科人群，应考虑其他儿科人群与青少年之间疾病、药物药理学和治疗反应的相似性，以及是否存在知识缺口。

决定将青少年队列纳入成人试验假设了疾病、药物药理学和治疗反应在青少年和成人患者之间足够相似。因此，将青少年和成人纳入同一试验的目标应在外推概念的框架内加以规划。由于青少年和成人PK通常相似，因此可能无需额外的数据来确定青少年给药方案。在这种情况下，应仔细考虑低体重对青少年给药方案的影响。在安全窗较宽的情况下，当给予相同的成人推荐固定剂量时，与成人相比，体重较低的青少年中出现更高的暴露量可能是可以接受的。

如果疾病和治疗反应足够相似，则可以将青少年和成人人群合并至同一有效性分析中。需要仔细考虑对青少年亚组

进行单独分析的统计方法和统计方法，以便解决任何已确定的差异或不确定性。应谨慎解释此类亚组分析；任何仅根据探索性亚组分析得出的有效性（或缺乏有效性）的外推结论，其强度可能有限（见ICH E9）。

将青少年纳入成人试验可能会面临伦理和操作方面的挑战，例如：（1）风险和潜在获益平衡的可接受标准不同；（2）青少年是否应暴露于安慰剂对照（在成人试验中的可能更常使用）；（3）除需获得青少年同意外，还需获得其父母许可；（4）在青少年和成人人群中使用相同的主要终点和安全性评估；以及（5）对儿科专科研究中心的需求。如果面临这些挑战，可以考虑采用不同的试验设计，或者在开展成人试验时平行开展一个独立的青少年试验。此外，整体儿科研发项目可能存在其他挑战，例如当下排除了青少年，儿科研究者是否愿意参与后续仅限儿科的试验。然而，当疾病、药物药理学和治疗反应在青少年和成人受试者之间足够相似时，强烈建议将青少年纳入成人临床试验或在平行试验中进行研究。

4.1.2 建模和模拟方法

建模和模拟（M&S）是儿科药物研发的重要工具，因为它为解决伦理约束、数据缺口、管理挑战、剂量优化，以及加速为儿科人群开发安全有效的治疗方法提供了手段。此外，其还有助于循证决策，从而在减少实验性试验带来的风险的

同时，为儿科患者带来更好的结局（见ICHE11R1）。与标准的M&S实践一致，应基于预期目的的充分性和适用性建立和评价模型。

建模与模拟方法可用于儿科外推，例如，评估疾病和治疗反应的相似性、检查和指导研究设计、制定给药方案建议、检验假设以及预测药物在目标人群中的作用。相关关系（如剂量-暴露量、E-R）的定量为进行模拟以支持剂量选择提供了重要基础。此外，在进行儿科研究之前，可以尝试与相关PK或PK/PD终点的治疗窗的模拟。在完成儿科研究后，建模和模拟可用于验证儿科外推概念所依据的假设。当模拟用于监管决策时，很重要的是要提供信息证明模型适合于模拟目的，并清楚地报告模型假设、输入数据质量和模拟设置。这些信息将以申办方为内部记录目的而生成的建模和模拟计划的形式提供，或者以与监管机构进行互动的适当报告形式提供。

各种数据源的可用性在一定程度上决定了建模方法，其中更具经验性的方法（如个体PK/PD、群体PK和PK/PD）依赖于参照人群的数据，而机制性方法（如PBPK、定量系统药理学[QSP]）依赖于诸如物理化学、体外和临床前体内数据等现有知识。当使用现有模型（如群体PK、PBPK、群体PK/PD模型）时，应将目标人群的特定特征（例如，相关体型、器官成熟度和其他需要的相关特征）纳入模型。根据可用数据

和建模目标，有几种技术可用于将参照人群的信息纳入目标人群的分析中；例如，使用基于参照人群的模型、使用汇总数据集进行分析，或采用具有模型参数先验分布的贝叶斯方法（见第4.2.3.4节）。在选择适当的技术时，应仔细考虑所选技术的优缺点和局限性。例如，在汇总目标人群和参照人群的数据时，重要的是确保目标人群的数据具有足够的代表性。否则，可能会导致参数估计受到参照人群数据的过度影响，并无法检测参照人群和目标人群之间的潜在差异。

在进行基于模型的评估时，模型的各个组成部分可能具有复杂的相互关系（如参数和/或假设的相关性），这些相互关系以及任何时间依赖性均应该在模型结构中得以体现。需要清晰地呈现模型结构所依据的模型方程和假设或数据集，以便能够适当地评估它们与总体策略、模型预测和不确定性因素的相关性。假设检验应整合到模型分析中。

区分不同来源的变异性和不确定性是十分重要的。例如，在个体之间采集的样本存在固有变异性（即受试者间变异性），这种固有变异性是一种生物现象，其程度可以直接得到数据支持。此外，由于数据不完整或对生物或生理过程（即模型假设）的理解不完整，模型参数可能存在不确定性。应说明并合理解释这些不确定性来源的不同贡献。应提供估计参数不确定性的程序。通常可以通过纳入更多信息数据来降低参数不确定性。鉴于模型假设和不确定性的程度，应该进

行多学科投入，以全面评估M&S结果和后续决策。

4.1.3 剂量选择

在适用的儿科亚组中评估和选择适当的待研究剂量对于实现目标暴露量和反应至关重要。在启动儿科研究之前，应评估关于药物作用机制、药物的药代动力学（ADME）、对预测药物和/或其活性代谢产物暴露量和反应可能产生的影响任何器官生理成熟程度和靶点，以及任何额外相关临床数据的可用科学信息（见第3.5节）。作为剂量选择计划的一部分，还应纳入其他考虑因素（例如：安全性、剂型、给药方案）。

从参照人群中收集的数据中得出的暴露量-效应关系可以为证明目标暴露量范围的合理性提供强有力的药理学基础。可结合相关知识和可用模型进行后续模拟，指导剂量选择（见第4.1.2节）。在参照人群中确定安全和有效的剂量，不一定总是需要或触发对于暴露量-效应(E-R)关系的证明。因此，在儿科人群中并没有一定要建立E-R关系的要求。即使在参照人群中无法证明暴露量-效应关系和/或无法证明参照人群和目标人群之间存在相似的暴露量-效应关系的情况下，仍然可以使用暴露量匹配。在将儿科患者随机分配接受亚治疗剂量或安慰剂可能不符合伦理要求，和/或现有安全性数据可能不支持更高剂量/暴露量相关的评估的情况下，在目标儿科人群中生成E-R曲线是不合适的。在这些情况下，基于暴露

量匹配和基于达到目标药物暴露量可能实现类似反应的预期的剂量选择是合理且务实的。

儿科剂量选择的目的是达到与参照人群中已知安全和有效的暴露量相似的暴露量，并在儿科有效性/安全性研究中进一步评估(见第4.2.3节)。为了确认选定的儿科剂量，通常需要目标人群的PK数据；然而，并非总是需要单独进行PK研究。可使用稀疏PK策略收集PK数据，作为儿科有效性/安全性研究的一部分。当达到目标暴露量所需的拟定剂量存在不确定性时(例如，发育和成熟变化可能会影响PK)，可能需要在临床研究中进行先导期PK评估以评价剂量的充分性。在某些情况下(如治疗窗窄的药物、非线性PK和/或参照人群与目标人群之间疾病对药物PK影响存在潜在差异)，应考虑单独进行PK研究。

另外，参照人群中的数据可能足以使用M&S(例如，群体PK、基于生理的药动学模型(PBPK)或其他建模和模拟方法)预测目标人群中的剂量。例如，如果存在来自不同儿科人群/适应症的试验药物暴露量数据，且这些儿科人群/适应症的年龄和暴露量范围与目标人群/适应症相同，则可能不再需要目标儿科人群中的PK数据。如在患有相同疾病的成人参照人群中观察到的暴露量数据可获得，并且目标暴露量在患有不同疾病的相同年龄的不同儿科人群中观察到的暴露量范围内，则可能不一定需要目标人群的额外PK数据。然而，

这种方法有赖于理解疾病对药物PK作用的影响。合理性依据充分时，即便观察到的药物暴露量超过了目标范围，仍可能通过建模和模拟，有足够的儿科PK数据来确定一个适宜的儿科剂量。

4.1.4 剂量范围数据的使用

作为儿科外推计划的一部分时，可能需要剂量范围数据。此类情况可能包括：当疾病相似性和/或治疗反应存在不确定性时；当靶点表达存在潜在的年龄相关差异时；或者当全身药物暴露量与治疗反应之间缺乏相关性时（例如：局部作用药物）。E-R和/或剂量-反应（D-R）关系可依赖于临床终点或生物标志物反应。根据生物标志物和疾病时程，可考虑通过剂量范围来获得不同程度的生物标志物/临床反应，或通过患者自身剂量滴定来达到目标生物标志物效应。

4.1.5 生物标志物的使用

生物标志物可在不同情况下使用，如在剂量选择中作为替代终点和/或作为桥接生物标志物（见第4.1.6节）。在可行的情况下，最好使用可同时支持成人和儿科研发项目的生物标志物。在相关情况下，作为成人研发项目的一部分，评估儿科外推计划中使用的潜在生物标志物可能是明智的选择。描述生物标志物与疾病进展和/或治疗效应之间关系的生理和/或机制性表征是非常有益的。M&S方法可用于支持儿科患者的生物标志物开发策略和临床终点的选择。建模和模拟

方法对于儿科患者中的生物标志物研发策略和临床终点选择可能非常有用。将生物标志物作为儿科外推计划的一部分应考虑以下因素：

- 使用经验证的生物标志物可能需要较少的合理性证明。建议使用经验证的生物标志物作为替代终点（非强制性）。

- 生物标志物终点的选择应受到参照人群和目标人群中可用数据的支持，并在外推计划中证明其合理性。

- 在参照人群中，与临床有效性相关的因果通路上的生物标志物通常是可接受的，应对其与目标人群的相关性也进行合理性说明

- 模型可用于估计生物标志物和临床有效性之间的定量关系（见第4.1.6节）。

与拟定生物标志物的稳健性和可靠性相关的方法学考虑因素（例如：缺失数据的影响、敏感性分析和对任何假设的偏离）应该得到解决。

4.1.6 建立参照人群和目标人群之间不同终点的关系

在创建儿科外推计划时，可能存在与可用于支持参照人群和目标人群之间的有效性的终点测量相关的差异。例如，参照成人人群中具有临床意义的终点（如6分钟步行试验行走距离）不适用于尚无行走能力的儿科患者。在这种情况下，应评估参照人群中使用的终点与目标儿科人群中的候选终点之间的关系。以下是确定用于目标人群的潜在终点的考虑

因素：

- 参照人群和目标人群之间是否存在相似的复合终点的组分？

- 参照人群中是否存在可作为目标人群的主要终点的次要终点？

- 是否存在类似的终点测量方法可用于参照人群和目标人群？

- 是否存在与参照人群中临床反应终点相关的生物标志物，这些生物标志物也与目标人群中的临床反应相关（即桥接生物标志物）

目标人群中选定终点的可接受性应基于现有证据的强度。当目标人群的拟定终点与参照人群不同时，尽早与监管机构讨论可能是有用的。

4.1.7 安全性外推计划

如第3.4.1节（安全性外推）所述，外推概念应包括对目标人群中安全性外推的讨论，并提供充分的理由来支持任何有关从参照人群向目标人群可以外推安全信息的结论的可接受性。儿科外推计划中描述的安全性数据收集方法应反映需要回答的科学问题、确定的知识缺口，正在解决的不确定性问题以支持药物在目标人群中的安全性。即使安全性外推是合理的，也可能存在一些选择性的儿科安全性问题，这些问题应在安全性外推计划中得到解决（见 ICHE19）。此外，

如果安全性窗较宽，目标暴露量高于成人暴露量可能是可以接受的。在安全性窗不够宽的情况下，靶向比成人更高的暴露水平，或者针对预测暴露量接受更高层次的不确定性可能就会出现问题。在某些情况下，除了作为有效性外推方法的一部分，在目标儿科人群中已经收集的数据之外，无需收集额外的安全性数据。如果确信已收集的可用安全性数据充分，并可解决相关安全性问题，则无需在儿科许可前项目中收集额外的安全性数据。

最后，应根据外推概念证明安全性外推的具体方法（包括上市前和上市后安全性数据收集的潜在需求）的合理性，并酌情与监管机构进行讨论。

4.2 儿科外推计划研究设计方法

方法的范围可能涉及：例如从参照人群中的有效和安全的暴露量匹配到在目标人群中生成同期对照的有效性和安全性数据。下列方法将会被讨论：

- 暴露量匹配方法（见第4.2.1节）
- PK/PD方法（见第4.2.2节）
- 有效性研究（见第4.2.3节），包括：
 - 单臂研究
 - 外部对照研究
 - 同期对照研究

此外，下文还讨论了研究设计、时机、分析、解释和报

告和/或儿科外推计划中包括的分析。

由于基于知识缺口和不确定性，外推概念可被视为具有演变连续性，因此有不止一种研究设计适合实现外推计划的目标。例如，单臂PK/PD研究的设计与依赖于临床有效性终点的单臂非对照研究之间可能存在一些重叠。此外，外推计划可以包括一个仅要求在目标人群中收集PK作为主要目标的情境，但也可以额外包含一些次要临床终点指标检测以增加“仅PK”方法的置信度。最后，应根据外推概念证明任何外推计划中使用的特定研究设计的合理性，并酌情与监管机构进行讨论。

4.2.1 暴露量匹配方法

当有强有力的证据 1) 支持参照人群和目标人群之间的疾病相似性时； 2) 参照人群中的暴露量将在目标人群中获得相似反应（如传染病、部分性癫痫发作）时，将参照人群中的有效暴露量作为儿科外推的基础（即暴露量匹配）可能是合理的。应采用M&S策略来支持目标人群PK研究中的初始剂量选择（见第4.1.2节）。异速放大方法可用于解释基于体重的清除率和分布容积的变化，并保持不同年龄/体重组的暴露量一致。模型还应考虑可能导致暴露量出现变化的其他因素，如成熟程度（见第4.1.2节）。此外，基于模型的剂量选择应包括对给药策略的可行性和实用性进行的评估。例如，固定剂量组合、剂量体积限制和药械组合可能会影响给药策

略。一旦获得目标人群的PK数据，应通过M&S技术对拟定给药方案进行评估。如果拟定的剂量方案未达到预期暴露量，则可潜在的使用M&S，使用现有数据推导出符合暴露量匹配标准的调整后的儿科给药方案策略，而无需在目标人群中收集额外的PK数据。

目标暴露量指标和暴露量范围

当儿科外推策略依赖于匹配在参照人群中的暴露量时，应预先规定目标暴露量指标、范围和可接受标准，并应在疾病、作用机制、治疗方案、给药途径和剂型的背景下对其进行定义。选定的目标暴露量指标应与参照人群的治疗反应相关，并且在安全性和有效性方面可以有所不同。例如，AUC或C_{min}可能与有效性相关，而C_{max}可能提供更多安全性信息。目标暴露量范围随后将根据在参照人群中已建立的暴露量-效应关系或观测数据得出。应根据（但不限于）作用机制和既往参照人群暴露量-效应关系中建立的指标，提供对拟定的指标及范围的充分讨论和合理性依据。多呈现几种暴露指标通常是有用的。在全身暴露量与有效性不相关的情况下（例如：大多数局部起效药物），可能需要对反应进行额外的评估。

样本量

儿科PK研究的拟定样本量应足以满足研究目的，并以定量方法（M&S和/或统计方法）为基础。应考虑并证明亚

组（如体重范围、年龄范围）具有充分代表性。样本量的合理性依据及其在目标适应症和年龄队列中的可行性评估应包括以下内容：

- 特定体重/年龄范围内患者的可及性
- 其他疾病人群的儿科PK数据的可及性
- 样本量是否足以证明儿科人群中关键PK参数（如清除率和分布容积）的准确度
- 用于确定样本量的方法

应采用建模和模拟技术（如最佳设计和/或临床试验模拟），以确定适宜的PK样本采样时机和数量。采集样本的时机和数量应尽可能与临床治疗保持一致（见ICHE11（R1））。

分析和报告

分析中应评估是否与在参照人群中预先指定的目标暴露量范围实现了匹配。应包括参照人群和目标人群中暴露量数据的不同呈现方式。针对所有药物和药物类别设定单一的可接受界值（与生物等效性检验相比）在儿科外推方面不会提供有意义的方法。应对关键暴露量指标（如AUC、C_{max}、C_{min}）的暴露量比和置信区间进行评估。所选择的暴露量之比的界值和置信区间显著性水平，应反映药物治疗范围的背景以及特定儿科适应症下产品的风险-获益。

应对参照人群和目标儿科人群中观察到的暴露量进行描述性比较。然而，许多儿科项目依赖于小样本量和/或稀疏

的PK样本，因此除描述性比较外，通常首选基于模型的方法。此外，应对整合所有相关观测数据的模型得出的暴露量指标进行比较（如适用）。另外，在确定暴露量相似性时需要考虑个体间变异性，而非仅仅比较平均值。对处于预定义暴露量范围内（或之外）的目标人群中不同年龄/体重范围内的受试者百分比进行模拟，可提供更有意义的暴露量相似性评估。应讨论观察/模拟数据与目标暴露量范围之间的差异。

一般来说，影响儿科患者PK的最相关因素为体重。在最年幼的儿科患者（如新生儿和婴儿）中，除体重外，还应考虑涉及药物处置的发育和成熟因素。相关的预先定义的暴露量指标，应使用连续或分类刻度，以与体重和/或年龄的关系图形式呈现。应以图表和表格形式呈现相关年龄和体重范围，以便清楚地显示重要的协变量（如剂量、年龄、体重）。还应以图表和表格形式呈现成人人群的参考范围（如观察或模拟数据分布的中位数和外百分位数）。

4.2.2 PK/PD方法

当仅靠暴露量匹配不足以确定有效性时，可将PD生物标志物作为外推计划的一部分。

为了依靠剂量/暴露量来实现生物标志物效应，确信参照人群中的生物标志物效应与有效性之间的存在关系非常重要。模型可以研究选定生物标志物的机制基础，促进生物标志物数据的分析，优化所需的数据收集，以支持和/或确认参

照人群中生物标志物和有效性之间存在关系（见第4.1.5节和第4.1.6节）。在外推计划中应定义生物标志物效应的治疗范围，以提供对参照和目标人群之间反应相似性的有意义的评估。

样本量

一般而言，第4.2.1节中描述的样本量考虑因素（在此也同样）适用。此外，应使用定量方法（M&S或统计方法）得出PK和/或生物标志物终点对应的样本量。研究的样本量可能因PK和生物标志物的变异性而有所不同。应考虑每个受试者的PK和生物标志物的数据点的采集时间和数量，以确定适当的采样。

分析和报告

应描述分析中使用的数据，重点关注与分析目的相关的重要元素，如目标人群和参照人群的生物标志物效应的比较。

应以适当的图表和表格形式对结果进行总结，如用于临床解释的说明图。应讨论结果的临床相关性，包括任何敏感性分析的影响。分析和报告应确认在目标人群中产生预期生物标志物效应的剂量/暴露量。

4.2.3 有效性研究

在多数情况下，需要生成有效性数据。有效性研究可以解决不同的科学问题，包括新的治疗是否对有效性产生有益的因果作用，以及目标人群和参照人群中治疗对结局的影响

是否相似。根据外推概念以及所确定的知识缺口和不确定性来决定最合适的目标问题。感兴趣的问题不同，选择的方法（包括频率论方法和贝叶斯方法）也随之不同。

根据ICHE9 (R1)，无论选择何种设计，旨在证明有效性的临床研究均应预先设定估计目标。由于参照和目标人群之间存在潜在差异，导致二者估计目标的某些属性可能无法完全一致（例如：给药方案、变量（终点）、伴发事件类型等）。当属性不同时，估计目标也会不同，此时采用相同的策略处理伴发事件可能不可行。但是，对于不同人群中的相同属性，建议尽可能对伴发事件采用相同的策略和分析方法。

当将外部数据纳入临床试验分析时，估计目标框架也可能有助于评估感兴趣的治疗效果。为确保对治疗效果进行有效地比较和可靠地估计，应仔细考虑因入组儿科人群与历史对照人群之间的差异（如患者特征、疾病进展）和伴发事件发生率的差异而引入的混杂因素及潜在偏倚。

研究的设计将取决于外推概念中确定的知识缺口。最重要的设计决策之一将是对照组的选择。选项可能包括单臂试验中可以与之正式比较的可以证明数据质量和相关性的外部对照组，或随机同期对照组。选择将受到儿科外推概念中确定的科学问题的影响。对于根据贝叶斯方法设计的试验，还有其他几个选择。本文件并非旨在提供关于模型选择的规范性建议，而是在儿科外推计划中设计有效性研究时提供重

要考虑因素。

4.2.3.1 单臂有效性研究

上文讨论了在外推计划中对目标儿科人群进行单臂PK研究的情况（见第4.2.1节和第4.2.2节）。然而，单臂有效性研究也可能是生成所需有效性证据最适宜的方法。可能出现这种情况的情形包括但不限于：在目标人群中缺乏合适的对照组或者在参照人群中获得批准的可接受证据即是单臂试验。

在设计此类研究时，应使用预先规定的标准（如成功的阈值或预先规定的准确度）来定义如何评估主要有效性目标。应在外推概念中确定阈值和准确度，并利用临床判断来确定可容忍的不确定性水平。

应计算研究的样本量，以便有足够的检验效能来确保达到成功的阈值，或确保获得足够准确度的估计值。可以使用外部数据（如已发表的文献或足够的可用真实世界数据源，如电子病历、索赔数据库或登记册）将结果与当前临床实践联系起来，但不需要针对有效性与外部数据进行正式比较。

4.2.3.2 外部对照研究

在某些情况下，使用外部数据作为试验中的正式对照是可能且适当的。这些数据可能来自于参照人群中的对照组、目标人群中其他随机对照试验（RCTs）中的相关现有对照组或目标人群中的真实世界数据（如来自观察性研究的结果）。

使用外部数据，如来自不同儿科人群、不同疾病或使用不同终点的数据，会导致更大的不确定性，并且可能需要额外的合理性证明。

与其他没有进行随机同期对照的研究一样，得出因果关系的推断更具挑战性。由于直接将数据与研究外部的数据来源进行比较，因此应使用适当的设计和统计方法来考虑人群之间的差异，尽一切可能减少影响结果的偏倚和混杂因素。重要的是要反映出，尽管采用了非随机对照，但这些研究仍为对照研究，这与仅与一个阈值进行比较的方法不同。

4.2.3.3 同期对照有效性研究

在某些情况下，迄今为止生成的数据和儿科外推概念的输出需要将随机对照有效性研究作为儿科外推计划的一部分，以便能够得出获益风险结论。对于对照研究的需求和外推能力通常导致研究设计与参照人群中所要求的设计不同。这将导致可接受的假阳性率、假阴性率和样本量之间的关系与参照人群中的不同。当样本量有限时，假阳性和假阴性错误率的相对重要性可以根据惯例（如双侧 p 值 <0.05 ）和参照人群中可能被认为合适的情况进行修改。如果采用信息先验的贝叶斯设计，则无法严格控制I类错误率。应根据具体情况仔细考虑，这两种错误风险之间的可接受折衷方案，并酌情与监管机构进行讨论。当根据治疗效果的特定准确度确定样本量时，错误率就会变得不那么相关。

由此可见，外推方案可能包括多种可用于生成数据的不同设计方案，但不一定遵循传统方法（例如：频率论分析方法中，RCT需要满足双侧 p 值 <0.05 ，才能判断试验成功）。外推法通常会导致样本量比独立有效性研究预期的样本量要小。如果研究的检验效能足够且被设计为满足放宽的成功标准（显著性阈值 >0.05 ），则应提前证明这种调整后的频率论方法的合理性。

设计阳性对照试验的另一种替代方法可能是保持传统的I类错误率，但扩大通常用于初治成人研发的非劣效（NI）界值，尤其是当目的不是证明有效性本身，而是证明有效性符合基于外推概念的先验预期时。或者，可以使用更宽的置信区间。

无论采用何种方法来证明非劣效，重要的是要确保治疗效果的点估计值不会引起劣效的担忧。

4.2.3.4 纳入外部数据

当确定应将哪些信息纳入儿科研究分析时，应通过适当的预先规定的方法（例如：使用预先规定的选择标准进行系统性文献综述）来确定相关数据。如有可能，应与监管机构商定好待使用的信息来源。然而，有可能外部数据本身目前还可能不可及，例如，外部数据是在参照人群中，与目标人群研究平行开展的试验中生成的，或者是从同一研究中不同的年龄组借用的。

分析中可以利用的信息类型包括来自其他来源的个体患者数据和/或汇总数据。在参照人群中有个体患者数据的情况下，就可以将基线预后因素的分布与目标人群进行比较。在分析中，应调整并说明得出参照数据的研究与目标人群中生成的数据之间的潜在差异。贝叶斯和/或频率学方法可用于合并来自参照人群和目标人群的数据，根据参照人群和目标人群之间的相似性评估来对参照数据的贡献进行加权。借用方法通常分为两类。第一类方法提前评估相似性，因此借用程度是预先指定的，不会根据观察到的数据而变化。另一种方法是动态借用方法，即预先指定一个模型，根据参照和目标人群之间观察到的基线特征和/或结局数据之间的一致性来调整借用程度。贝叶斯和频率学中都有动态借用方法存在。

4.2.3.5 量化使用参照数据的影响

利用频率学方法使用试验外额外的数据，能够清晰地了解借用了多少数据。相比之下，使用贝叶斯方法可能就不够清晰。当使用信息贝叶斯先验时，重要的是应事先了解有多少可用信息被纳入分析中以支持对儿科试验的解释。尤其相关的是要知道两类独立的信息：我们预计在训练中将使用多少参照人群的信息，即基于参照数据的先验的有效样本量（ESS）；其次，相对于使用的参照信息的先验ESS，将在目标人群中产生多少数据。对于一些贝叶斯模型，先验ESS，将在目标人群中产生多少数据。对于某些贝叶斯模型，有多种

方法可用于估计ESS。ESS可能是固定的，也可能会根据观察到的数据和所选择的模型而变化。在这种情况下，应记录下ESS的范围。如果可用信息（基于参照数据或建模和模拟训练的输出）总结为统计分布，那么ESS则是描述使用了多少信息的好方法。

如果使用贝叶斯方法，并且通过的不同方式使用了先验信息，例如通过使用混合先验（例如，由信息和弱信息成分混合组成的先验）或幂先验，将根据所做的假设和在先验构建中选择使用的参数产生不同的ESS。如果采用了此策略，尤其是在先验数据存在冲突的情况下，则可以通过进行敏感性分析来查看这些参数在不同取值情况下的ESS，从而更有助于了解设计特性。无论使用哪种方法，都应预先规定拟使用的借用方法，用于了解不同借用量对操作特性影响的敏感性分析，将进一步辅助了解设计特性。

当已知参考人群和目标人群之间存在疾病差异时（例如，疾病严重程度），如果这些差异可以量化和调整，则外推概念仍然适用。如果这些差异导致无法直接使用参照数据，则应对数据进行建模，以更加准确地预测目标人群中的有效性。在其他情况下，尽管疾病被认为相似程度足以进行外推，但是研究设计可能存在已知差异（例如，在目标人群中测量的终点不同，或者在不同的时间点测量终点）。此时，如何使用参照数据，将根据疾病、药物药理学和治疗反应的相似程

度具体情况具体分析。

有可能在目标人群中将生物标志物、替代终点或临床终点作为主要终点来制定儿科外推计划，即使在参照人群中它并不是主要终点[见ICHE11(R1)和第4.1.6节]。在这种情境下，应对拟定终点与参照人群中主要有效性终点的相关性是否稳健进行评估。此时，可能明智的做法是在将其纳入儿科项目之前，作为成人开发计划的一部分，启动对潜在儿科终点的评估。

4.2.4 儿科试验的呈现和合理性依据

可呈现整体计划的试验设计的图表是很有帮助的，尤其是当设计复杂时。这种情况可能出现在，例如存在适应性设计或有多个阶段，每个阶段评估临床研发的不同方面的试验中。在评估试验设计时，根据预先定义的标准确定哪些潜在结果将促使研究成功，有助于了解需要观察到何种治疗效果程度才可宣布试验成功。如果最合适的阈值存在不确定性，则不同临界阈值的表格或图表可能是有用的。

应对研究设计进行评估，包括在与计划假设不一致的情况下（例如：存在先验数据冲突的情况）。当使用贝叶斯设计（包括稳健的混合先验）时，这一点尤其重要。无论选择何种设计（以及是否使用频率学方法或贝叶斯方法），都应在评估中确定设计的操作特征（例如：假阳性和假阴性错误率）、估计量的特性（偏差、方差）和区间特性（例如：频

率学的置信区间或贝叶斯的可信区间)。此外,应始终提供仅来自目标人群的数据的频率学分析结果。

4.2.5 分析、报告和解释

与预先规定的估计目标一致的准确分析、全面的报告和对结果的明确解释对于确保得出可靠结论和做出明智决策至关重要。如果使用频率学设计,除标准双侧显著性水平5%外,还可以采用替代阈值,应证明其合理性并进行预先规定。如果外推概念支持适宜将参照和目标数据一起正式分析,则可以实施合并参照和目标数据的频率学荟萃分析方法。

如果使用贝叶斯设计,明确利用外部数据,则需要为分析做出更多选择。虽然先验信息可能会随着额外的外部数据产生而更新,但这种分析应该是预先指定的。可视化工具有助于更好地理解操作特性与基本参数和假设之间的关系。后验分布图可以更好地将贝叶斯分析产生的汇总统计信息置于上下文中理解。如果将试验外部的数据纳入分析,则应在报告中明确说明这一点,并讨论这些数据最初生成的方式和时间,申办方从什么来源获取用于分析的数据,并合理说明它们为何被认为适合纳入分析。

无论是从贝叶斯还是频率学的角度来看,数据是否足够相似,使得能够按照外推计划中提出的程度进行合并,取决于目标人群中生成的证据。理想情况下,如果事先与监管机构讨论并商定了成功标准(如适用),将有助于研究的解释。

可以使用的成功标准可以是p值，或者如果明确借用了参照数据，则可以是贝叶斯成功标准，例如不包括临界值的可信区间，或者一种治疗优于另一种治疗至少某一预定量的概率。可能可以适用不止一个成功标准。例如，如果使用的非劣效界值比成人可接受的范围更宽，则还可能可以指定治疗效果的点估计，需要证明对于任何给定的样本量和方差都满足非劣效性。这有助于通过为预期治疗效果提供额外的保证来证明有效性。这有助于通过为预期治疗效果提供额外的保证来证明有效性。

如果研究中的观察数据偏离了观察到的参考数据，这可能会限制儿科外推概念的适用性，以及可能被视为合理借用的数据量。然而，尽管如此，如果目标人群中的数据在效应点估计值方面明显优于参照人群，但由于样本量小而未能实现在未借用的情况下达到成功标准，那么在做出积极结论之前（即使用临界点分析），了解需要给予这些参照数据多少权重可能是有意义的。

在某些情况下，对目标人群中生成的数据的一种解释可能是，这些数据是否与根据外推概念所预期的数据一致。在这种情况下，研究成功的标准和评估方法应能反映研究目标。对数据的这种解释需要事先证明所得出估计值的准确性和由此产生的样本量的合理性。

统计模型越复杂，需要假设的参数越多，就越需要适当

和更广泛的敏感性分析[ICH E9(R1)]。事先讨论这些敏感性分析，并研究主要分析的解释对这些参数变化的稳健性程度是有益的。应仔细选择此类分析，以研究使用主要估计量所做的假设和数据的其他限制。

4.2.6 在儿科试验分析中利用参照数据的方法

需要证明分析中使用的参考数据的选择是合理的。在决定使用何种方法时，模拟作为一种有用的工具，可为分析策略的选择提供指导，以优化偏倚、检验效能（假阴性）和I类错误率（假阳性）之间的权衡。如果生成的数据与其相关的先验信念不相似，则存在各种旨在限制借用的方法。例如，在众多方法中，一种可能的方法是使用稳健的混合先验：例如双分量混合先验，其中一个分量是基于参照数据的信息先验，第二个分量是独立于参照证据的弱信息先验。应谨慎选择弱信息分量，因为如果指定的方差过大，可能会导致对极端、临床不合理的参数值给予过大的权重，从而影响预期的借用行为。归因于混合先验信息分量的先验权重可被视为关于外推概念合理性和可接受性的先验信念。该值越接近1，置信度越高。如果预先规定的参数（如上述加权参数）发生微小变化，导致研究的操作特征发生较大变化，则该方法可能不够稳健。

敏感性分析是回顾性评估关于先验假设强度结论稳健性的有用工具。其中一个例子是临界点分析，评估了贝叶斯

先验中指定参数值的变化，以查看有效性结论是否存在变化。

当参照数据来自多个不同来源时，如成人RCTs、流行病学研究或登记注册数据，不同来源的数据质量可能不同，其与新的儿科试验的相关性也可能不同。在这种情况下，应仔细考虑先验本身的构建以及将数据纳入分析的方法。