

临床试验中的估计目标与敏感性分析

E9(R1)

当前版本：第2阶段草稿

2017年6月16日

临床试验中的估计目标与敏感性分析

E9 (R1)

目录

1. 目的与范围.....	1
2. 规划、设计、实施、分析和解释的框架.....	3
3. 估计目标.....	4
3.1 描述.....	4
3.2 解决伴发事件的策略.....	5
3.3 估计目标的构建.....	7
3.3.1 一般考虑.....	7
3.3.2 治疗和试验背景的考虑.....	8
4. 对试验设计和实施的影响.....	10
5. 对试验分析的影响.....	11
5.1 主要估计.....	11
5.2 敏感性分析.....	12
5.2.1 敏感性分析的作用.....	12
5.2.2 敏感性分析方法的选择.....	12
5.3 补充分析.....	13
6. 记录估计目标和敏感性分析.....	14
7. 通用示例.....	14
7.1 一种伴发事件的情形.....	15
7.2 两种伴发事件的情形.....	19
词汇表.....	21

1. 目的与范围

准确知情患者或医生以帮助正确选择药物，需要有对药物效应清晰的描述。患者个体对治疗反应的不同使得这些描述变得复杂。对同一种药物，有些患者可以耐受，并能遵守用药时间表，而有些则不能；一些患者需要改变药物剂量，或使用合并用药，有些则不需要。定量评价治疗效应的方法有多种，可以基于耐受性、依从性以及是否需要合并用药等分别加以考虑。若没有对治疗效应的准确理解，药物效应的大小及意义就存在被误解的风险。

通过验证性临床试验，通常是随机对照试验，我们得以量化治疗效应，并提供有效性与安全性证据，以支持监管决策。随机试验一般不会出现基线混杂，但在临床试验中会发生某些事件，使疗效的描述和解释变得复杂。在本增补中，伴发事件（intercurrent event，见术语表）指试验治疗开始后出现的事件，包括使用其他治疗（如补救治疗，方案违禁用药，或后线治疗），停止治疗，治疗转组（switching），或出现终端事件（如在某些情况下的死亡）。

选择和定义有效性和安全性变量、数据收集标准，及统计分析方法时，若没有事先考虑伴发事件的发生，可能会导致待估计的疗效模糊不清，甚至背离试验目的。正确的顺序应是反过来的，即在试验设计阶段描述待估疗效时，需明确试验目的，仔细考虑伴发事件，并说明试验设计、数据收集和统计分析方法的选择。

本增补提出了一个结构化框架，将试验目标与适当的试验设计，及估计和假设检验的工具联系起来。该框架介绍了估计目标（estimand，见词汇表）的概念，将试验目的转化为对待估治疗效应的准确定义（第A.3节），旨在促进涉及临床试验设计、实施、分析和解读的多学科之间，以及申办方与监管者之间，就临床试验所评价的治疗效应进行对话交流。估计目标相对应的统计分析假设和数据局限性相关，其影响可通过敏感性分析（sensitivity analysis，见词汇表）进行探查和评价。本增补阐明了敏感性分析的定义和作用。ICH E9中的引用使用符号x.y，本增补中的引用使用符号A.x.y。

本增补在若干方面澄清并拓展了ICH E9。

首先，ICH E9介绍了意向性治疗（intention-to-treat, ITT）原则，其与疗法策略（treatment policy）的疗效相关，该策略的疗效指根据基线初始分配的治疗组相对应的疗效，且无论是否依从，疗法都保持随机化，这为统计学检验提供了一个坚实的基础。随机分组是临床对照试验的基石，应尽可能地利用随机分组的优势进行分析，这是无可争议的。然而问题在于，依据疗法策略得到的效应是否总与监管和临床决策所关注的治疗效应最一致？该增补中给出的框架为讨论其他治疗效应提供了一

个基础，并提出了设计和分析时需考虑的要点，以便提供治疗的效应估计值，为决策提供可靠依据。

其次，本增补重新审视了数据处理和缺失数据（missing data，见词汇表）的问题。一方面，诸如终止治疗、治疗转组或使用补救治疗之类的伴发事件，即使后续测量可以收集，在某些情况下也许与研究无关或难以解释。或在死亡的情况下，受试者死亡后的测量不存在。另一方面，ICH E9注意到当临床试验对象停止治疗后失访，很难遵循ITT原则。本增补考虑了不依从或退出随机分配的治疗与终止试验的重要区别，以及存在但尚未收集的测量值与根本不存在的测量值之间的区别。明确估计目标，为收集数据的计划提供了基础，计划哪些数据需要收集，而哪些数据在未被收集时会出现缺失数据问题，并选择与估计目标相一致的缺失数据处理方法。

第三，在提出的框架中考虑了分析集的概念。（ICH E9）第5.2节对优效性试验的分析强烈推荐全分析集（full analysis set, FAS），即尽可能接近包括所有被随机化分组的受试者。然而，试验中通常会收集同一个体的重复测量值，剔除认为无关紧要或难以解释的某些个体测量值，这种做法与从全分析集中完全剔除一个个体一样，会产生同样的后果，即未能完全维护初始的随机化。此外，如果受试者已经死亡，尽管结果变量的值会有意义，但并不存在。（ICH E9）第5.2节并未直接涉及这些问题，而要使这些问题更加清晰，需仔细界定待估计的疗效，确定纳入疗效评估的受试者人群，将每个受试者的观察结果纳入分析，并且考虑可能的伴发事件发生。符合方案集（per-protocol analysis, PPS）的意义和作用也在本增补中进行了重新审视；对于是否需要评价方案偏离或违背的影响，通过某种方式，比简单进行符合方案集的分析，偏倚更少且更具有可解释性。

最后，在敏感性分析的标题下，进一步讨论了稳健性的概念。特别指出了在一个在特定假设下特定分析的推断敏感性，与更广泛地选择分析方法的敏感性的区别。通过准确指定估计目标和相应的统计方法，分析的细节要达到能使第三方精确地重现分析结果，这样，监管的重点可以放在对假设偏离和数据局限性等方面的特定敏感性分析上。

2. 规划、设计、实施、分析和解释的框架

为提高一致性，使研究问题更清晰，试验计划应按顺序进行（图1）。通过定义适当的估计目标，将明确的试验目标转化为某个关注的科学问题。估计目标通过指定目标人群、目标变量、伴发事件的处理以及目标人群层面的效应量（第A.3节）来定义。一个估计目标（*estimand*）对特定的试验目的（即“要估计什么”）定义了估计的目标（*target of estimation*）。然后可以选择合适的估计方法（即分析方法，称为主估计量）。主估计量（*main estimator*）将以某些假设为基础。为了探索主估计量推断的稳健性，针对同一估计目标（第A.5节），应以一个或多个分析的形式进行敏感性分析。

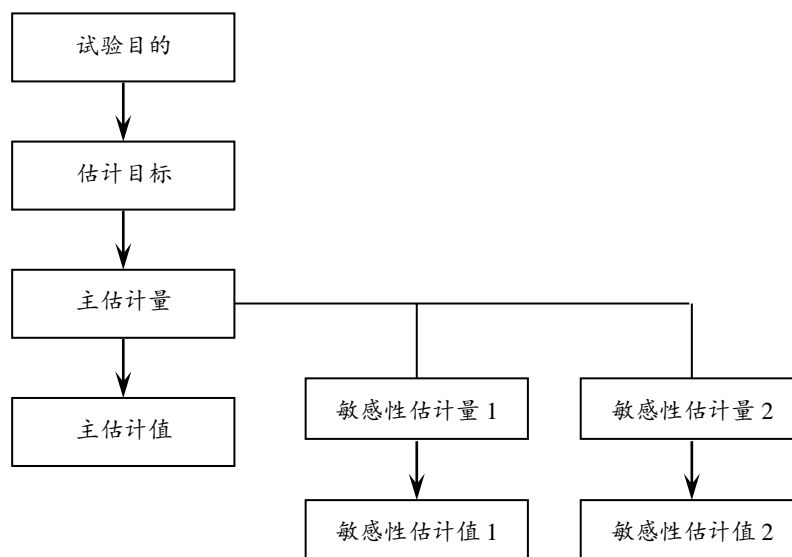


图1. 对给定的试验目标，制定相一致的估计目标、估计方法和敏感性分析

该框架能够进行适当的试验计划，从而明确区分估计的目标（试验目的，估计目标），估计方法（估计量，及由此得到的相应的估计值，见术语表），和敏感性分析。该框架有助于申办者进行试验设计，监管机构进行审查，并在讨论临床试验设计的适用性和解释临床试验结果时，促进申办者和监管机构之间的交流，以支持药物的上市许可。

一般来说，根据框架按序进行非常重要，由此阐明待解决的科学问题，而不是先选择估计量再确定估计目标。

指定适当的估计目标（见A.3.3）通常是试验设计、实施（A.4节）和分析（A.5节）等方面的主要决定因素。

3. 估计目标

3.1 描述

药物开发和许可的核心问题是量化治疗效应，即相同受试者其治疗效应与其他治疗（如与未接受治疗或接受其他治疗相比）相比结果如何。在对感兴趣的变量所反映的治疗效应进行描述时，需要考虑伴发事件，因为变量取值和事件是否发生可能与治疗有关。通过指定一个估计目标，应该考虑疗效的定义与伴发事件之后的变量取值是否相关，以及如何考虑事件是否发生本身（可能与治疗相关）。

更正式地，对某一具体的科学问题，估计目标（*estimand*）详细定义了需要估计什么。一个估计目标包括四个属性：

- A. 目标人群（*population*），即科学问题所针对的患者；
- B. 目标变量（或终点），取自每个受试者，用于阐述待解决的科学问题；
- C. 说明如何处理伴发事件，以反映感兴趣的科学问题；
- D. 目标人群层面的效应量，为不同治疗条件下提供相比较的基础。

这些属性共同描述了估计目标，定义了待估计的疗效。

在大多数情况下，目标人群是指根据方案中的纳入/排除标准、有资格被纳入临床试验的受试者。在某些情况下，这些受试者的某一层（*stratum*）可能是有意义的，例如，自始至终配合治疗的受试者这一层，这个层通常以潜在的伴发事件来划分。

目标变量通常包括测量（例如血压测量）、测量的函数（例如，HbA1c从基线到一年的变化），或与临床结果相关的数量（例如，死亡时间，住院次数，复发次数）。目标变量还可以同时考虑伴发事件（如停止治疗），例如，使用在停药之前进行的测量（如，HbA1c曲线下的面积直到停止；在治疗期间血压得到控制的周数），或综合变量（例如，将无应答和/或终止治疗定义为治疗失败）。

有必要明确指出如何考虑潜在的伴发事件才能更好的解答感兴趣的科学问题。伴发事件可以以多种形式出现，并可影响目标变量的解释。例如，若受试者在计划的血压测量时间点之前死亡，则不会观察到该时间点上的血压。如果受试者除了治疗之外还使用了补救治疗，此时虽然能观察到血压，但这反映了治疗药物和补救治疗的综合效应。如果受试者由于毒性而终止治疗，则也能观察到

血压,但这反映了在未用药时的情况。考虑的一系列伴发事件将取决于具体的治疗情境和试验目的。以补救治疗为例,可有两种不同的考虑,即处理与任何伴发事件(在这种情况下使用补救治疗)的综合效应,以及未发生伴发事件时的(理论上的)治疗效应。A.3.2节描述了在构建估计目标的过程中解决伴发事件的不同策略,以使其与相应的科学问题相一致。

第四个属性是目标变量在目标人群层面的效应量,例如,定义目标变量为从HbA1c从基线至一年末的平均变化,或满足指定应答标准的受试者比例。在两个治疗组比较的情况下,综合评估分别为:两组HbA1c从基线到一年末的变化值均数之差值,或两组满足指定应答标准的受试者比例之差或之比。

3.2 解决伴发事件的策略

A.3.1节介绍的估计目标的属性A到D是相互关联的,不应该被单独考虑。如果在科学问题中没能反映伴发事件的影响,那么这个估计目标的描述就是不完整的。解决伴发事件至少可以考虑五个策略。这些策略可以单独使用,也可以联合使用以解决多种不同的伴发事件,与估计目标的其他属性一起,选择解决伴发事件的策略,针对性地描述治疗效应。A.7节提供了一些示例,介绍了在考虑一个或多个伴发事件时,如何使用这五种策略来构建估计目标。

每个策略的相关性将取决于治疗和试验背景。此外,在每个试验情况下,使用这些被认为可靠的决策策略构建的特定的估计目标,有可能不能导出一个估计值。这些考虑在A.3.3, A.3.4, A.4和A.5节中讨论。以下所列仅供参考,在构建估计目标时必须对所选的策略给予适当描述。

疗法策略 (treatment policy strategy)

伴发事件的发生与估计目标不相关,即无论是否发生伴发事件,均采用目标变量的值。

例如,当指定补救治疗为伴发事件时,忽略伴发事件的发生,并使用目标变量的观察值。如果对所有类型的伴发事件均按此处理,则是ICH E9术语表ITT原则中描述的比较疗法策略的效应。

一般而言,当个体伴发事件之后变量的值不存在时,无法实现该策略。例如,在出现死亡病例时,目标变量无法测量,基于此策略的估计目标就无法构建。

组合策略 (composite strategy)

伴发事件的发生被认为是目标变量的一个组成部分,即伴发事件与临床结果的一个或多个其他测量结果整合为目标变量。

在该策略下有多种方法可以考虑。补救治疗的需求可能提供了有关治疗效应的有用信息，因此可以将其以适当的形式整合到一个变量中，用以描述治疗效应。例如，该变量可以定义为一个复合变量，即未使用补救治疗并出现有利的临床结果；再如，当目标变量是数值变量时，将发生的伴发事件用一个极端值表示，以示最差的情况，并选择某种合适的方法进行分析。另一种方法是使用曲线下面积（AUC）反映计划的随访时间内、但在伴发事件之前发生的变量值。

有时，伴发事件本身被认为是衡量治疗效应的最有意义的变量。死亡就是这种情况，受试者死亡的事实可能比死亡前的观察值更有价值，而死亡后的观察值不会存在。例如，在主要关注心肌梗死的试验中，不一定总能够确定死亡的受试者是否患有心肌梗塞，但是如果将该结局变量定义为一个复合事件，如死亡或发生心肌梗死，就能够得到确定的结局。

假想策略（hypothetical strategy）

设想一种不会发生伴发事件的场景：反映科学问题的测量值定义为在假设场景中的变量值。

例如，当出于伦理原因必须提供补救治疗，而待评价的疗效是未使用补救治疗时的结果。类似地，在治疗失败和停止治疗后使用另一种阳性治疗（包括事先随机分配到试验组的受试者转组到对照组中），但待评价的疗效是没有使用随后的阳性治疗。在这些例子中，未提供补救治疗和未采用阳性治疗反映了不同的假设条件。

在具体试验的背景下需要准确地描述反映科学问题的假设条件。例如，假设条件可能有助于将使用补救治疗和治疗依从性作为伴发事件，以便对估计目标进行准确描述。

主层策略（principal stratum strategy）

目标人群可定义为不会发生伴发事件的主层（principal stratum，见词汇表）。例如，将不会发生不依从的受试者层定义为目标人群。换句话说，主层是不会发生伴发事件的较广泛人群的一个子集。科学问题仅限于在该主层的治疗效果研究。

任何类型的亚组分析或符合方案分析的受试者组成基于试验数据，值得注意的是，主层效应与此有明确的区分，主分层（principal stratification，见词汇表）由受试者在两种治疗中的潜在伴发事件定义，例如，会依从某种疗法的受试者。一般来说，因为伴发事件的发生不能预测，这些患者无法在试验之前直接识别；他们也无法基于随机对照试验的数据来识别，因为每个受试者只在一种治疗中。因此只能通过协变量推断出主层个体，尽管这种做法仍存在瑕疵。相比之下，基于指定治疗组的伴发事件得到的主层个体，进行治疗效应评估，容易产生混淆，因为不同的受试者将在不同的

治疗组经历不同的伴发事件。

在治策略 (while on treatment strategy)

在治策略感兴趣的是在伴发事件发生之前的治疗效果。如果是重复测量变量，变量取值是伴发事件发生前的所有测量值，而不是所有受试者在相同的固定时间点的值。

例如，患有终末期疾病的受试者可能因为死亡而终止一种单纯症状治疗，但治疗是否成功可以基于死亡前症状的改善效果来测量。或者，受试者可能会终止治疗，在某些情况下，评估在治疗依从期内不良反应的风险也是有意义的。

总而言之，本节考虑了五种不同的策略。描述处理每个伴发事件的首选策略时要准确到位。考虑治疗依从性时，区分下列不同情况下的疗效是至关重要的：(i) 假设情景“若所有受试者都会依从”，(ii) “若在试验组治疗就能够依从”的受试者层，以及(iii) 依从期间的效果。

3.3 估计目标的构建

3.3.1 一般考虑

如上所述，为了准确描述待估计的治疗效果，并使治疗效果的描述对于患者和医师更为相关，必须在构建估计目标时明确考虑伴发事件。估计目标的构建应解决临床试验中可能发生的每个伴发事件，这将影响对试验结果的解释。在设计阶段，对伴发事件的描述理论上应能反映治疗和随访的非常具体的细节，例如指定变量观察的具体时间窗。预期这些具体标准不会影响试验结果的解释。要预见所有相关伴发事件是不切实际的。但试验报告应讨论如何处理未预见的伴发事件，同时讨论对所选的分析估计值的影响。在构建一个估计目标时，可以选择不同的策略（见A.3.2, A.7节）来处理不同的伴发事件。

任何一个临床试验中，估计目标的构建是一项多学科的工作，包括临床医生、统计师和其他涉及临床试验设计和实施的学科。也是申办者与监管机构就前瞻性临床试验目的及设计进行交流讨论的重点。估计目标的构建应该是试验目的的结果，并提供有关数据收集和分析方法的选择考量。回避或过度简化此过程，会导致试验目的、试验设计、数据收集与统计分析彼此间的不一致。

估计目标的构建过程可能是需要反复进行的。考虑在特定治疗情境（包括研究的疾病、治疗目标以及特定试验情境（A.3.3.2节））中的临床相关性，应合理构建估计目标。此外，需要考虑试验设计和统计方法的充分性，以确保得到一个推断可靠的估计值。特别是随机化在临床试验中的关键

优势应尽可能地得到承认和利用。一些估计目标，特别是那些使用观察数据估计的估计值，只有很少的假设，可以得到稳健估计；而其他估计则需要更多具体的假设，这些假设可能难以证明，并且可能对这些假设条件的变化更敏感（见A.5.1节）。如果对一个特定的估计目标，不存在合适的试验设计，且难以导出合理的估计值，则需要考虑备选的估计目标、试验设计和分析方法。

3.3.2 治疗和试验背景的考虑

如上所述，疾病种类、治疗目标将影响估计目标的构建。基于具体治疗内容，这可能包括：备选治疗的可行性、监测个体对治疗反应的可能性，以及治疗目的是针对控制症状、缓解疾病进程还是预防疾病。例如，治疗的目标可以是控制临床症状或体征，而在该疾病领域存在多种备选治疗，并且可以根据观察到的反应来定制受试者的治疗选择，此时使用备选治疗（补救治疗，方案禁忌用药，或后线治疗）可能需要被视为伴发事件。若无备选治疗，或者可以坚持治疗而不需要备选治疗的受试者层，说明如何处理此类伴发事件需要基于对该治疗效果的了解以反映相对应的科学问题。由于疗法策略不能区分受试者是否接受了其他治疗，在某些情况下，是否接受其他治疗的答案可能比量化疗法策略的疗效更为相关。若受试者从分配的对照组转组到试验组，这类伴发事件显得尤为相关。若能获得稳健的估计值，可以使用这些策略构建估计目标。在其他情况下，有必要了解疗法策略在临床实践背景下的治疗效果。例如，治疗的目的可以是预防或延缓不利的临床结局（例如，死亡）。在后线的疗法已确立的情况下，若治疗被预设用于初治受试者，疗法策略的效果会得到更大的关注。当基于疗法策略构建估计目标时，可以通过定义与伴发事件本身相关的额外的估计和分析来补充推论，例如，以症状评分测量的治疗效果，与在各种治疗条件下使用补救治疗的数量进行对比。

有些情况下，虽然基于其他策略的估计可能被认为有更大的临床兴趣，但对主要估计量和敏感性估计量是否能够提供可靠的估计或稳健的推断无法达成共识，这时，基于疗法策略的估计可能会被更普遍地接受，以支持监管决策。基于疗法策略的估计目标，可提高获得相关治疗效果的可靠估计的可能性。在这种情况下，建议保留那些被认为具有更大临床意义的估计目标，提供所得到的估计值，并讨论对该方法的试验设计或统计分析方面的限制。

复合策略的一个例子是用二分类变量替换连续变量，在未发生伴发事件的情况下，可以预设变化值的阈值，将受试者分为应答和未应答（有效和无效）。因此，把连续性变量转化为二分类变量将引起估计目标的变化。估计目标的临床相关性和解释将取决于应答的标准和适当的目标人群效应量（例如，比例之差，优势比）是否可用。

对于假想策略，某些情况下可能比其他更适合于监管决策。因此，采用假想策略所描述的假设条件必须进行调整，使得解释的治疗效果的量化与临床实践中该药物的使用相关联。如上所述，如果无法使用补救治疗，针对这一条件下的治疗效果，比起包括了使用补救治疗的疗法策略，目标变量取什么样的值，就会变得重要。相比之下，在假想策略的条件下，因药物不良反应，理论上应停止治疗，但实际上并没有停止而是继续治疗的受试者，此时得到的变量值，无论是从科学性还是监管部门角度可能都会被认为是不合理的。如果假设所有受试者会依从，若没有对这一假想条件进行充分论证，感兴趣的对于效果的科学问题就无法很好地定义。进一步的，在试验和临床实践中无法耐受的治疗本身，则可能很难取得有利结果的证据。如果处理伴发事件的选择的策略，不仅取决于例如缺乏依从情况，还取决于缺乏依从的原因（如毒性），则必须定义这些事件，并准确记录在临床试验中。

另外还应该考虑试验情况本身。如果某临床试验方案下的受试者管理（例如，因耐受性而调整剂量，因受试者不适而采取补救治疗等）被证明与临床实践中预期的不同，则应该反映在估计目标的构建中。特别值得一提的是，对照组的选择可能影响试验中补救治疗的使用，及试验中其他伴随用药。

使用除所分配治疗以外的其他治疗通常被认为是伴发事件。其他治疗可以是多种多样的，包括补救治疗、方案中的违禁用药，或提前使用后线（subsequent line）治疗。而且即使补救治疗也有多种形式，包括对缺乏疗效的受试者使用、替换或者增加一种慢性治疗，当受试者无法耐受试验药物时替换成其他治疗，或者增加短期应急的对症治疗。这些例子说明了在特殊实验情况下考虑特定伴发事件处理的重要性。

对于非劣效性或等效性的试验，研究的估计目标需要仔细斟酌。在（ICH E9）3.3.2节中指出，这种试验本质上并不保守，尽量减少试验方案违背和偏离、不依从和试验撤离的发生是非常重要的。在（ICH E9）第5.2.1节中，描述了这类试验中全分析集（FAS）的结果通常不保守，其作用应该审慎考虑。考虑一个或多个伴发事件采用疗法策略而构建的估计目标，对于非劣效性和等效性试验有类似的问题，这些考虑与使用FAS的问题是类似的。因为与最初分配治疗间的相似性无关，在终止所随机分配的治疗或使用另一种药物后，两个治疗组的反应将更趋于相似。估计目标的构建则可以直接解决可能导致治疗间差异变小的伴发事件（例如使用补救治疗，和目标人群对方案的违背）。在这种情况下，估计目标可选择较高灵敏度的指标（如果存在的话）来检测治疗组间的差异。

4. 对试验设计和实施的影响

试验的设计需要与估计目标的选择相一致，以反映主要试验目的，在此基础上来确定试验目的是否达成。具体来说，在明确定义估计目标时，需要说明选择何种与试验设计一致的疗效量化方法。例如，如果兴趣在于了解无论是否发生特定伴发事件下的疗效，合适的试验则应收集所有受试者的变量，而不管该伴发事件是否发生。或者，如果用于监管决策所需要的估计目标不需要在伴发事件之后收集变量，那么则需要把为其他估计目标而收集变量的好处，与收集的复杂性和潜在的弊端进行权衡。

为进行与感兴趣的估计目标及其他重要的估计目标相一致的统计分析，应努力收集所有相关数据以支持分析。与终端事件（terminal event）（如死亡）不同，诸如不依从、终止治疗、治疗转组或使用补救治疗等伴发事件的发生，并不意味着此后不能测量该变量。如果不收集评估估计目标所需的所有数据，将导致后续的统计推断面临缺失数据问题。取决于缺失数据比例，统计分析的有效性会受到不可检验的假设的影响，从而可能会破坏结果稳健性（第A.5节）。如果有计划的收集数据缺失的原因，将有助于区分关注的伴发事件和其他缺失数据，从而潜在的改善主要分析。这也会对选择敏感性分析有帮助。例如，或许通用的“失访”应正确记录为“由于缺乏疗效而终止治疗”。通过选择合适的处理伴发事件的策略，那些可以被伴发事件定义的就不是一个缺失数据的问题。需要采取一些措施保留受试者在试验中，但也应注意考虑保留临床实践的外部有效性(external validity)。例如，选择试验人群，使用滴定方案或伴随药物来减轻毒性的影响，若相同的措施在临床实践中无法实施，这些措施则是不合适的。

某些估计目标采用非标准试验设计可能更有益，例如有导入期（run-in）的设计，富集设计（enrichment design），随机撤离设计（randomised withdrawal design）或滴定设计（titration design）。然而，这些设计可能需要特别考虑实施和后续的统计推断。例如，在将受试者随机分配到治疗组和对照组之前，尝试使用导入期（run-in period）来确定受试者可耐受的水平。监管机构与申办方之间需要沟通考虑导入期是否适合来定位目标人群，以及为后续试验设计的选择（例如，清洗期，随机化）是否支持来估计预定治疗效果和相关推论。这些考虑可能会限制这些试验设计和特定的策略在实践中的使用。

通过指定处理伴发事件的策略，对待评价的疗效的准确描述与样本量计算相关。如果所有个体对分析提供信息，并且伴发事件的影响及其处理反映在目标效应大小和预期方差上，通常不需要将算得的样本量按预期的撤离（withdrawal）比例进行放大。

第7.2节讨论了多个临床试验间数据总结的相关问题。目标变量有一致的定义需要被强调，这拓展到建立估计目标。因此，欲将多个临床试验数据进行汇总，应在设计阶段计划相应的程序，构建一个合适的估计目标，包括试验方案，并反映在所选试验（contributing trials）的设计选择上。类似的考虑适用于荟萃分析（meta-analysis）的设计，或解释单臂试验中的外部对照组。若未考虑也未指定估计目标，简单比较数据来源或整合分析多个试验数据，则会引起误导从而产生偏倚。

更普遍的情况，一个试验可能将多个试验目的转化为多个估计目标。某种试验设计可能适用于某种估计目标，但可能不适合其他可能重要的估计目标。具有多个目的和多个终点指标的试验，需要考虑多重性检验（multiple testing）的问题，原则上，对多个估计目标也同样需要有多重性考虑。特别是涉及一个或多个主要估计目标而另一些是次要估计目标时，多重性检验的方法同样适用于多个估计目标的情况。

5. 对试验分析的影响

5.1 主要估计

治疗效果的估计目标应能反映治疗组中受试者相对于对照组中相类似的受试者的临床结局（outcomes），从而使治疗效果可以从治疗组间差异中区分出来。对于给定的估计目标，需要有与之相对应的分析方法或估计量，以便得到估计值，并能给出可靠解释。能否提供稳健的估计，很大程度上取决于假设的条件。所做的假设应与主估计量和敏感性估计量一起明确说明。假设应该是合理的、切合实际的。在假设条件下结果的稳健性应通过与估计目标相一致的敏感性分析进行评估（第A.5.2节）。

特别需要说明，无论是否发生伴发事件，如果完成了受试者完全的随访，基于疗法策略的估计则只需要最少的假设。如果在伴发事件之后未收集测量值，则该策略下估计目标的估计将需要更强和不可检验的假设。在原始变量尚未完全确定的情况下，使用组合策略，有可能不需要在伴发事件之后对结果进行填补或建模进行分析，也不需要相应的假设。相反，如果解决伴发事件基于某假设情境，构建估计目标的估计时需要仔细说明假设条件，并且必然依赖于建模假设，而模型的假设条件是否成立是无法检验的，需要通过敏感性分析来进行评价。在随机试验中，除非在随机分组之前即可鉴定主层内的受试者，否则目标人群在主层的疗效估计是存在混杂的。或者估计将依赖于假设，且需要所有相关的混杂因素被测量且被适当方法考虑。例如，对于预期能依从的受试者层，若简单地比较治疗组的依从者和对照组的依从者是不恰当的。这些受试者存在着系统偏差，从而混杂治疗

效果的估计。在这种情况下，解释所有重要的混杂因素就非常的重要，而不仅是某个小的、先入为主的协变量集，此时也难保证能避免错误的指定模型。对于在治策略，当伴发事件的发生和发生时间与治疗有关时，治疗效果的估计将需要更强的假设。

在定义估计目标时，尽管已经以适当的方式解决伴发事件，并努力收集估计所需的数据（第A.4节），一些数据仍然可能缺失。这种缺失数据与收集估计所需信息的系统失败或空缺不同。例如，如果构建了一个基于疗法策略的估计目标，则应尽全力使受试者留在试验中，即使在停止所分配治疗后，受试者仍能依从随访时间表进行访视。在这些努力失败时，有必要对缺失数据进行假设，基于观察到的数据对个体缺失值进行预测或填补，或调整统计方法。缺失数据的处理应基于合理的假设，并在可能的情况下以估计目标的描述中所采用的策略为指导。某个体的预测则基于该受试者的（协变量和基线后）数据或其他相似个体观察到的数据，这些相似个体的标准可能包括是否已经评估了伴发事件（例如，对于停止治疗但未进一步收集数据的受试者，预测模型可能会使用停止治疗但进一步收集了数据的其他受试者，而不是坚持治疗的受试者）。与这些技术方法的假设存在合理的偏差，是敏感性分析的一个重要方面。

5.2 敏感性分析

5.2.1 敏感性分析的作用

即使数据有局限性，以及与主估计量的统计模型假设偏离时，基于某估计目标的统计推断应当是稳健的。这种稳健性常用敏感性分析来评估。

支持主要估计目标的统计假设应该记录在案，针对同一估计目标，需事先确定一个或多个分析以对假设进行探讨，目的是验证从主估计量到估计值，对于假设偏离是否稳健。与敏感性分析不同，为更全面地研究和了解试验数据，计划、呈现或要求的其他分析均称为补充分析（supplementary analysis，见词汇表）。每个补充分析可以针对不同的估计目标，或同一估计目标下的不同估计量。这里，主要估计目标是经申办者和监管机构同意的，主要估计量也是事先明确规定的。敏感性分析通常比补充分析更重要，应优先考虑敏感性分析。

5.2.2 敏感性分析方法的选择

在计划和实施敏感性分析时，建议不要同时考虑改变过多的条件，否则，当出现不一致的结果时，难以确定是哪个因素的影响（如果有的话）。一个明显且实用的方法是一次只改变一个假设条

件。此外，在评估不同分析结果的解释和相关性时，区分可检验和不可检验的假设条件会有帮助。

在敏感性分析中需要特别注意缺失数据，因为任何方法的假设可能难以证明完全合理，也可能根本无法验证。缺失数据必须在特定的估计目标下定义和考虑。例如，基于疗法策略的估计目标，在终止试验药物之后仍应收集的数据没有收集，则算是缺失；然而，相同数据则可能与另一个策略无关，对于相应的次要估计目标，即使未收集，也不算缺失。庆幸的是，相关类型的假设偏离通常可以简单地表征。例如，在分析连续性结果变量的均值时，因每个治疗组的缺失和非缺失数据在均值上的差异，原始分析可能产生不同程度的偏倚，特别是这些差异存在治疗组间差异时。此时应研究这些差异假设值的合理范围，并对结论的稳健性进行评估。例如，在假设检验中，在指定的检验水准下，治疗组间有统计学意义或没有统计学意义的差值，可以在临界点分析(tipping point analysis)中画出。也可以根据达到或未达到具有临床意义的差值，来评价疗效之差。类似的技术可以应用于其他数据结构。例如，可以假定对于成功率或生存资料的风险，在缺失数据和非缺失数据之间是不同的，且在治疗组间也是不同的。

5.3 补充分析

如果经敏感性分析已验证相应的估计值是稳健的，试验结果的解释应重点关注每个已经确定的估计目标的主要估计量。

尽管针对不同估计目标的补充分析可以提供额外的见解，但对解释试验结果仅仅是起了次要作用。例如，应答率的分析比基于连续变量的平均差值的量化结果更容易解释。此外，对应答的不同定义可用于探查结果是否对该定义稳健。对每个试验都应确定补充分析的必要性及其作用。

(ICH E9) 第5.2.3节表明，同时基于FAS和符合方案集(PPS)进行分析通常是合适的，因为它们之间的差异可以成为明确讨论和解释的主体。如果FAS和PPS的分析结果一致，表明对试验结果更有信心。同样在(ICH E9) 第5.2.2节中指出，基于PPS的结果可能会存在严重的偏倚。根据本增补中提出的框架，可以进行基于被指定到某特定治疗组且试验方案依从性好的受试者子集的分析，但分析本身并不一定能明确地界定感兴趣的治疗效应。如上所述，基于符合方案集的分析并不能达到估计疗效的目标，因为不同治疗组间的受试者可能不尽相似。因此，这种分析的作用仅限于探究方案违背和偏离的程度是否会降低对试验结果的信心。一些方案偏离和偏差可作为伴发事件处理。通过估计目标的建立可处理大多数伴发事件，所剩其他的方案偏离和偏差者数量将会很少，由此PPS的分析可能不会提出新的见解。

6. 记录估计目标和敏感性分析

估计目标应在临床试验方案中明确规定。试验方案中应该事先确定与试验目的相对应的主要估计目标，描述那些可预见的伴发事件，及其对试验结果解释的影响。此外，试验方案和分析计划应事先确定与主要估计目标相一致的主要估计量及主要分析，以及适当的敏感性分析，以探讨主要分析在偏离假设时的稳健性。同时也应该适当描述可以支持监管决策的次要试验目的（例如与次要变量相关的）的估计目标，以及相应主估计量和适当的敏感性分析。其他试验目的可作为探索性分析，也需进一步定义估计目标。

尽管明晰所要估计的内容对申办方是有利的，但监管并不要求对每一个探索性问题的估计目标进行详细的文档说明，特别在处理伴发事件时主要的和次要的估计目标只有微小变化。然而，如果不同的科学问题所需估计目标有本质的区别，那么建议将这些问题充分记录在案。

主要估计目标的选择通常是试验设计和实施方面的主要决定因素，按照惯例，这些应在试验方案中详细记录。如果对其他估计目标也特别感兴趣，可根据需要，考虑拓展试验的设计和实施，这些当然也应记录在案。除此之外，试验设计、实施和分析方面的常规考量仍然是一样的。例如，如果存在多个估计目标，并引起潜在的多重检验问题，则控制一类错误的通常考虑仍适用，并相应地进行描述（见A.4节）。

主要分析、敏感性分析和补充性分析的结果应在临床试验报告中系统地报告，并具体说明每项分析是在盲态下事先确定的，还是事后进行的。对于在设计阶段没有预见到的伴发事件，或在试验进行过程中才明确的伴发事件，不仅要讨论伴发事件的处理分析方法和估计值的选择，而且应讨论对试验结果解释的影响。

7. 通用示例

以下将以连续性变量作为通用示例说明本增补中提出的框架，该基本框架对二分类(binary)变量及时间事件(time to event)变量等同样适用。这些例子不应理解为监管建议，根据具体临床试验的情境，需要做出相应的调整。

对某位非危及生命的慢性疾病患者，考虑一种新的治疗（药物X）。使用连续性变量，每月监测对治疗的反应。药物X的效果预计在治疗开始后四至六个月可以看到。主要科学问题是药物X与安慰剂在治疗第6个月的效果比较，随机临床试验是回答这个科学问题的最佳选择。在该临床试验中使用

安慰剂符合伦理，只有在受试者缺乏疗效而终止治疗后才切换到补救治疗。补救治疗是一个伴发事件，之后仍然可以收集主要变量的测量值。对其他伴发事件也是如此，例如由于不良事件而终止治疗，死亡这类伴发事件除外（在这种情况下不可能继续收集主要变量的测量值）。

假设无伴发事件的发生（虽然通常并不现实），合理的估计目标需包括以下4个方面：

A.目标人群：定义恰当的纳入/排除标准，以确定受试者目标人群；

B.目标变量：指定测量指标从基线到6个月的变化值；

C.伴发事件：这里不需要考虑伴发事件；

D.目标人群层面的效应量：不同治疗组间均值之差。

因此，该试验的估计目标是目标人群从基线到第6个月的目标变量之变化值在不同组间的均值之差。

为此估计目标，采用随机对照平行设计，在整个试验中收集所有测量结果，未收集到的视为缺失数据。疗效指标收集完成后，以治疗组为影响因素的方差分析（ANOVA）模型可作为统计分析方法之一。缺失值填补方法需事先设定，需要根据由于缺失数据而导致的不确定性的合理假设来预测数据。例如，缺失值可用继续进行试验的相似受试者的测量值进行填补，这里，相似可定义为：相同的基线指标、同一治疗组、同一测量时点，以及伴发事件情况。敏感性分析需要在方案中事先指定，以评估缺失值填补方法的假设等问题。根据研究需要，可增加补充分析，例如治疗组间的转组比例和时间等。

这里简化属性C的标题为“伴发事件”，该部分主要描述如何考虑潜在的伴发事件，及如何反映在感兴趣的科学问题中。

7.1 一种伴发事件的情形

实际操作中，伴发事件时有发生。为简单起见，在试验中仅考虑出现“因缺乏疗效切换到补救治疗”的伴发事件。下文描述发生伴发事件时的估计目标和对应的科学问题，以及试验设计、实施和分析的进一步考虑。

疗法策略

A. 目标人群：定义恰当的入排标准，确定目标受试人群；

B. 目标变量：事先计划的基线至 6 个月的测量变化值；

C. 伴发事件：不管是否进行补救治疗；

D. 目标人群层面的效应量：不同治疗组间均值之差。

按照疗法策略，该案例估计目标中的效应实际为“药 X+补救治疗”与“安慰剂+补救治疗”相比较的结果。因此，估计目标中同时混杂了治疗效应和救助治疗效应，其影响大小取决于采用补救治疗者在两组所占的比例。此外，补救治疗是由于“缺乏疗效”而驱动的事件，这一估计目标并没有明确考虑这一信息。

这里的估计目标定义为：目标人群自基线至 6 个月的测量变化在不同治疗组间的均值之差，无论是否转组到补救治疗。类似语句也将用在下面每个例子中，同样整合关于如何处理伴发事件的说明。

采用这个估计目标的试验设计是随机对照平行设计，无论是否发生补救治疗，收集试验过程中所有治疗组疗效测量。

所有数据收集完成后，以治疗组为影响因素的方差分析模型可作为估计目标的统计分析方法之一。若存在缺失数据，需在合理的假设前提下进行估计。例如，缺失值可根据继续试验的相似受试者的数据进行填补，相似性可定义为：基线协变量相同、随机治疗组相同、相同的伴发事件测量历史和信息。敏感性分析需在试验方案中事先定义以评估缺失值填补方法的假设。针对估计目标，可进行补充分析，例如比较两组补救治疗的比例和发生的时间。也可进行补救治疗相关的其他感兴趣的估计目标分析。

组合策略(Composite strategy)

A. 目标人群：定义恰当的入排标准，确定目标受试人群；

B. 目标变量：在无补救治疗的伴发事件前提下，基于基线至 6 个月的效应变化值，根据事先定义的界值，将 6 个月随访时是否有疗效作为二分类疗效指标；

C. 伴发事件：目标变量定义中已做考虑；

D. 目标人群层面的效应量：不同治疗组间有应答比例之差。

上述估计目标的组合策略仅考虑了第 6 个月的疗效评价。治疗效应为组合变量(composite variable)，采用了是否有临床意义的二分类变量，以及转到补救治疗组的伴发事件。因转至补救治疗组是由于疗效缺乏，固此类估计目标认定进行补救治疗即是无效。

采用该估计目标的一个试验设计是随机平行组设计。除非有其他研究目的（例如收集伴发事件发生后的安全性信息），发生转组之后无需收集测量值。根据在此示例中的估计目标，转至补救治疗组后未收集的数据不属于缺失数据。

一旦确定反应状态的所有测量数据收集完成后，对此估计目标采用 logistic 回归模型进行统计分析是一个例子。如在伴发事件之前数据缺失，需要根据合理的假设进行填补。例如，缺失值可根据继续试验的相似受试者的数据进行填补，相似性可定义为：基线协变量相同、随机治疗组相同、测量历史相同。敏感性分析需在试验方案中事先定义以评估缺失值填补方法的假设。针对组合估计目标的某一部分可进行补充分析，例如，重新定义二分类变量界值后的估计目标的分析。

假想策略 (hypothetical strategy)

- A. 目标人群：定义恰当的入排标准，确定目标受试人群；
- B. 目标变量：事先定义的基线至 6 个月的测量变化值；
- C. 伴发事件：6 个月之前不发生补救治疗转组的伴发事件；
- D. 目标人群层面的效应量：组间的目标变量均值之差。

假想策略所描述的估计目标阐述患者未获得补救治疗的假想情境下治疗效应。直接针对该科学问题进行临床试验，可能存在伦理问题。

试验采用随机平行组设计。除非有其他研究目的（例如，收集伴发事件发生后的安全性信息），发生转组后的测量值无需收集。在这个示例中，因与欲估计的估计目标无关，转至补救治疗组后该收集而未收集的数据不作为缺失数据。

该估计目标的统计分析将基于受试者未接受补救治疗这一假设前提。通常，需要借助敏感性分析考察结论的稳健性。关于假设合理性的讨论，可提高假设的可靠性，以及疗效估计的可靠性。针对估计目标可进行补充分析，例如比较组间接受补救治疗的比例和时间。

主层策略 (principal stratum strategy)

- A. 目标人群：在满足入排标准的目标人群中，无论接受何种治疗，6 个月期间不要求补救治疗的受试者人群；
- B. 目标变量：基线到 6 个月的变化值；

C. 伴发事件：伴发事件通过人群定义所收集；

D. 目标人群层面的效应量：组间目标变量均值之差。

通过主层策略，这个估计目标评估的是对于无论哪个治疗组，6 个月期间不要求补救治疗的人群的治疗效应。

在实际操作过程中，此方法的复杂性在于难以事先定义这类人群。因此，可能需要采用非标准的设计，以保障目标人群在 6 个月期间不接受补救转组治疗，例如富集设计，有导入期的设计和随机撤离设计等。

该估计目标的统计分析只针对 6 个月内不发生补救转组治疗的随机化受试者进行。如上所述，由于很难事先确定这类受试者，目标人群将无法确定，分析亦无法进行。如果剔除发生补救治疗的受试者，则不同治疗组间可能存在系统差别，治疗效应将包含影响受试者转组的混杂因素，需要恰当的分析方法考虑混杂因素。另外，可以通过敏感性分析来评价分析结果的稳健性。针对不同的估计目标，还可进行补充分析，例如比较组间采用补救治疗的人群比例和时间。

在治策略 (while on treatment strategy)

A. 目标人群：定义恰当的入排标准，确定目标受试人群；

B. 目标变量：随机化治疗期间的平均测量值；

C. 伴发事件：已考虑在目标变量的定义中；

D. 目标人群层面的效应量：治疗组间目标变量均值之差。

该估计目标通过目标变量测量值的平均值来评估治疗效应。选择的目标变量采用转组补救治疗前所有疗效的平均值作为估计。

该估计目标的一个试验设计是采用随机平行组设计。除非基于其他研究目的（例如其他研究目的所需伴发事件发生后的安全性信息），转组后的测量值无需收集。对于此示例中的估计目标，转组后该收集但未收集的数据不属于缺失数据。

随机化治疗组的测量值收集完成后，以治疗组作为因素的方差分析模型可作为该估计目标的统计分析方法。对于间断的测量缺失值，可基于合理的假设对数据进行插值估计 (interpolated)。敏感性分析需在方案中事先定义，以评价插值估计方法的假设。针对不同的估计目标进行补充分析能够辅助推断，例如改变目标变量的定义，只关注治疗期间的最后一次测量，将导致不同的估计目标。

7.2 两种伴发事件的情形

现将示例扩展至可能发生两种类型伴发事件的情形，即“转组到补救治疗”和“因不良事件停止治疗”。临床上有意义的估计目标的定义需要包括临床试验中可能发生且与临床相关的所有伴发事件，如果估计目标中不包括所有伴发事件，则不能完全理解估计目标中的疗效。对设计、实施和统计分析的选择也应考虑所有伴发事件。上述五种策略，可以考虑所有可能的策略之组合以处理两种伴发事件，虽然不是所有的策略组合都临床相关。简单起见，以下仅描述了两种不同的估计目标的策略，以及关于试验设计、实施和分析的进一步考虑。

考虑两种伴发事件的疗法策略

- A. 目标人群：定义恰当的入排标准，确定目标受试人群；
- B. 目标变量：从基线至第六个月的变化值；
- C. 伴发事件：不管是否使用补救治疗，或因不良事件而终止；
- D. 目标人群层面的效应量：治疗组间目标变量均值之差。

这个估计目标针对治疗启动后基于变量观测的疗法效应。这个估计目标既不解释补救治疗的启用，也不解释因不良事件治疗终止。事实上，该策略并没有使用补救治疗和不良事件这些不利结果信息。

针对这一估计目标采用随机平行组设计，所有测量均在整个试验中收集，无论是否使用补救治疗，或因不良事件而终止。

只要收集所有的测量值，以治疗组为因素的方差模型分析是对该目标估计合适的统计分析。在测量缺失的情况下，需要根据由于缺失数据而导致的不确定性的合理假设来预测数据。例如，缺失值可根据继续试验的相似受试者的数据进行填补，相似性可定义为：基线协变量相同、随机治疗组相同、相同的伴发事件测量历史和信息。敏感性分析应在试验方案中进行事先确定，以评估填补方法的假设。可以通过包括额外的补充分析来补充推断，可能针对不同的估计目标，例如比较不同治疗组的补救治疗转组的比例和时间，以及因不良事件终止试验例数。

两种伴发事件的假想策略和疗法策略之组合

- A. 目标人群：定义恰当的入排标准，确定目标受试人群；
- B. 目标变量：从基线至第六个月的变化值；

C. 伴发事件：6个月之内未向受试者提供补救治疗，无论是否因不良事件而停止治疗；

D. 目标人群层面的效应量：组间目标变量均值之差。

这种估计目标结合了两种不同的策略来解释两种伴发事件。采用假想策略来处理补救治疗，采用治疗策略处理不良事件导致的治疗终止。这样的目标估计会引起关注，在以药物效应为目的的情境下容易解释，但该用而未用补救治疗不符合伦理，且对因不良事件而停止治疗后的受试者没有给予适当治疗。

针对这一估计目标的设计是随机平行组设计，无论是否因不良事件而终止治疗，所有测量均在整个试验中收集。在补救治疗后，不需要收集测量值，除非有其他试验目的需要这种数据。在这个示例的估计目标中，在使用补救治疗后，该收集而没有收集的数据不认为是缺失数据。

针对这种估计目标的统计分析需要考虑到两个伴发事件：

- 转组到补救治疗：感兴趣的是假设第6月前受试者未使用补救治疗情况下的治疗效应。由于这种情况无法直接测量，所以需要在这种假想情境下的测量进行假设。

- 因不良事件引起的治疗终止：感兴趣的是忽略这种伴发事件发生的治疗效应。因此，所有测量都需要包含在分析中，无论这种伴发事件是否发生。在测量缺失的情况下，需要基于合理的假设来预测数据，同时要考虑由于数据缺失而增加的不确定性。例如，缺失数据可能会根据试验中的相似受试者进行估算。相似性可定义为：相同的基线协变量，相同的随机分组，相同的伴发事件测量历史和信息，如发生时间等。

一旦个体预测与观察到的伴发事件和感兴趣的估计目标一致，使用基于所有随机对象的统计分析（例如方差分析模型）是合适的。在缺失数据情况下，需要基于合理的假设来预测数据，并考虑由于缺失数据而导致的不确定性。例如，缺失数据可能会根据试验中的相似受试者信息进行填补。相似性建立在如下基础上：相同的基线协变量，相同的随机分组，相同的伴发事件测量历史和信息。敏感性分析应在试验方案中进行事先确定，以评估填补方法的假设等。可以针对不同的估计目标，通过补充分析来完善统计推断，例如，比较组间补救治疗的比例和时间，因不良事件导致的治疗终止的比例和时间。

词汇表

估计目标 (Estimand) :

是为了阐述由试验目的提出的科学问题而进行估计的目标。一个估计目标包括四个方面的属性：目标人群、目标变量（或终点）、科学问题中如何反映伴发事件的说明，以及目标变量在目标人群层面的效应量。

估计值 (Estimate) :

基于观察到的临床试验数据，由估计量 (estimator) 计算得到的数值。

估计量 (Estimator) :

基于观察到的临床试验数据计算估计值 (estimate) 的分析方法。

伴发事件 (Intercurrent Events) :

在治疗开始后发生的事件，会影响目标变量的观察或影响其解释。

缺失数据 (Missing Data) :

对设定的估计目标的分析有意义，但未收集到的数据。必须把缺失数据与不存在的数据，或因伴发事件而被认为没有意义的的数据区分开来。

主分层 (Principal Stratification) :

是根据所有治疗的伴发事件潜在发生情况，对受试者进行的分类。比如对于两种治疗的试验，基于给定的伴发事件，有四个主要层次：在两组的任一组都不会经历伴发事件的受试者；接受治疗A会经历伴发事件者，但接受治疗B不会经历事件者；接受治疗B会经历伴发事件者，但接受治疗A不会经历事件者；受试者在两个治疗组均会经历伴发事件者。

主层 (Principal Stratum) :

在本文件中是指主分层定义的任何层（或层的组合）。

敏感性分析 (Sensitivity Analysis) :

是针对同一估计目标，在偏离基本建模假设和有数据局限性时，基于不同的假设来探索主要估计量推断结果的稳健性，而进行的一系列分析。

补充分析 (Supplementary Analysis) :

是指主要分析或敏感性分析之外的其他分析，以便更多地了解治疗效应。比敏感性分析更广泛的一类分析。