

国家药监局药审中心关于发布《患者报告结局在药物临床研发中应用的指导原则（试行）》的通告（2021 年第 62 号）

发布日期：20220104

为鼓励以患者为中心的新药研发理念，科学合理的运用患者报告结局，药审中心组织制定了《患者报告结局在药物临床研发中应用的指导原则（试行）》（见附件）。根据《国家药监局综合司关于印发药品技术指导原则发布程序的通知》（药监综药管〔2020〕9 号）要求，经国家药品监督管理局审查同意，现予发布，自发布之日起实施。

特此通告。

附件：患者报告结局在药物临床研发中应用的指导原则（试行）

国家药监局药审中心

2021 年 12 月 27 日

患者报告结局在药物临床研究中应用的 指导原则 (试行)

2021 年 12 月

目 录

一、 引言	1
二、 患者报告结局的定义	1
三、 患者报告结局测量量表的研发、翻译、改进	2
（一）患者报告结局测量量表的研发	2
（二）用于患者报告结局测量量表的翻译和/或文化调适	7
（三）患者报告结局测量量表的改进	9
四、 患者报告结局测量量表的选择与评价	9
五、 临床研究中使用患者报告结局的考虑	10
（一）估计目标框架	10
（二）选择患者报告结局作为临床研究终点	11
（三）研究方案和研究报告中有关量表的阐述	11
（四）量表的有效应答	12
（五）缺失数据	13
（六）多重性问题	13
（七）结果的解释	14
（八） PRO/ePRO 的质量控制	15
（九）真实世界研究中 PRO/ePRO 的使用	15
六、 电子化患者报告结局	16
（一） ePRO 测量	16
（二）使用 ePRO 的一般考虑	17
七、 与监管机构的沟通交流	18
参考文献	20
附录 1: 词汇表	26
附录 2: 中英文词汇对照	30

患者报告结局在药物临床研究中应用的指导原则

（试行）

一、 引言

临床结局是评价药物治疗获益与风险的核心依据，如何准确、可靠、完整地观测临床结局至关重要。患者报告结局 (patient-reported outcome, PRO) 是临床结局的形式之一，在药物注册临床研究中得到越来越广泛的使用。另外，随着患者为中心的药物研发 (patient-focused drug development, PFDD) 的理念和实践的不断发展，在药物全生命周期中获取患者体验、见解、需求等数据并将其有效地融入到药物的研发和评价中日益受到重视，临床报告结局 (clinical outcome assessments, COA) 特别是其中的患者报告结局可以反映患者的感受，是患者为中心的药物研发的重要组成部分。

本指导原则旨在阐明 PRO 的定义以及在药物注册研究中的适用范围，PRO 测量特别是量表研发和使用的一般原则，PRO 数据采集的质量控制，数据分析和解释需要注意的事项，以及与监管部门的沟通等，为申办者在药物注册研究中合理使用 PRO 数据提供指导性意见。

本指导原则适用于使用 PRO 作为终点指标支持药品注册的临床研究，包括临床试验和真实世界研究。

二、 患者报告结局的定义

患者报告结局定义为：任何来自患者直接报告且不被他人修改或解读的对自身疾病和相应治疗感受的评估结局。

PRO 强调患者自己报告结局，当患者不具备或丧失自我评估能力时，可能需要由其监护人或监护人指定的代表完成 PRO 的记录，但此时应充分评估代理人偏倚。

量表是 PRO 测量使用较多的工具，主要用于主观测量，如疼痛、生存质量等，但现有量表并不能解决所有的主观测量问题，如某些症状（如恶心）或症状群。PRO 的数据采集有纸质记录和电子化载体两种手段。使用电子化手段记录 PRO 称为电子化患者报告结局（electronic patient-reported outcome, ePRO）。

三、 患者报告结局测量量表的研发、翻译、改进

临床研究中，一旦确定使用量表测量 PRO，如果尚无适合研究项目的量表，需专门针对研究目的进行研发；如果已有公认的适合研究项目的中文量表，在获得版权后可直接使用；如果已有公认的适合研究项目的外文量表，需经过研发形成正式的中文版本后使用；如果已有量表并不完全适合研究项目时，需改进后使用。在已有的成熟量表中如何选择更适合于拟开展的研究项目，需要考虑其科学性和可操作性。

（一） 患者报告结局测量量表的研发

PRO 测量量表的研发应能反映患者的视角，重点考虑该量表的临床价值，包括疗效评价的针对性、临床意义的可解释性和对治疗决策的指导性。量表的研发过程如图 2 所示。量表的研发通常用于有效性评价，也可以针对重要的安全性事件进行研发，其原理和过程是一样的。

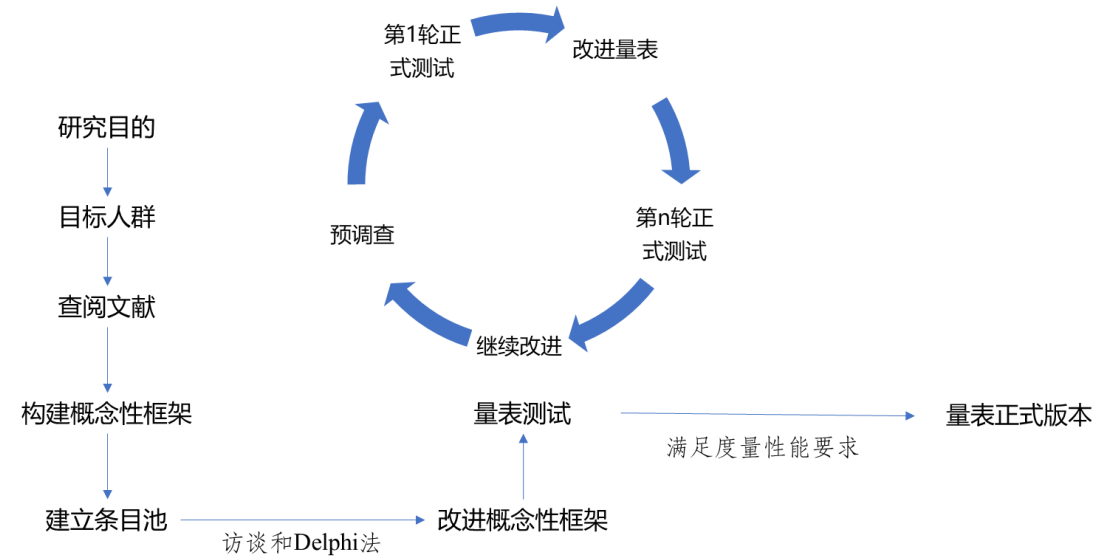


图 2 量表研发过程示意图

1. 构建概念性框架

量表的结构有一级结构、二级结构和三级结构，临床研究中以一级和二级结构较常用。一级结构的量表有单条目量表（如视觉模拟疼痛量表）和多条目量表（如简化版口腔干燥量表）。下面以二级结构量表为例进行阐述。

二级结构量表的第一级是维度，第二级是条目。量表概念性框架的初步成型一般基于研发者查阅文献、专家知识和经验、患者访谈以及必要的调研。维度的数量和命名依据对

研究内容的理解设定，每个维度下的条目数和条目内容用以体现其所属维度的内涵和重要程度，例如每个条目等权时，维度下的条目数量就体现了维度的重要性。

2. 建立条目池

量表的底层结构是条目，体现具体的设问内容，而维度则是概念性的。为了后续的条目设计，需建立尽可能丰富的条目池，条目的来源可以是所有可能的途径，包括文献、患者和/或专家访谈、相关领域的量表研发平台、研发报告、研发者设计等。

条目设计是量表研发的核心内容之一。如果条目池足够丰富和成熟，绝大多数的条目一般从条目池中获取，但也会有一些条目由研发者设计。在问题的陈述中，应尽可能采用封闭式问题，避免含混不清的词语、具有双重含义或倾向性引导的问题、双重否定的陈述、以及负面陈述和患者不情愿回答的问题；同时应避免应答的天花板或地板效应，以及一个条目同时问两个以上问题等。在阅读理解方面，尽量使用常用语，对文化水平的要求不宜太高（如具备小学毕业文化程度的阅读能力即可）。

3. 标度方法

条目的标度有二分类标度、等级标度（如 Likert 标度）、连续标度（如视觉模拟标度）、图形标度等方法，其中以 5

级 Likert 标度法最常用。具体采用几级 Likert 标度要以量表的度量性能达到最佳为标准。

4. 访谈

在研发者初步形成了量表的概念性框架后，首先需要进行患者访谈、专家访谈和/或专家调查，根据专家反馈意见调整概念性框架。患者访谈有助于进一步保证患者报告结局量表的内容有效性，体现患者的需求和意见。专家调查的主要目的是考证结构的合理性、条目表述的准确性、应答的可行性和维度及条目的赋权。维度和条目的赋权是量表研发最为关键的环节。专家调查法的实施通常不止一轮，以达到专家意见相对统一为止，特别是条目赋权方面的意见。

5. 预调查和正式调查

在综合专家意见改进初始的概念性框架后形成量表的初始测试版，继而需要在目标人群中进行测试，然后根据测试结果改进，形成正式测试版。使用正式测试版在目标人群中展开调查，其样本量需根据预调查的参数进行估计，正式测试版的改进也是根据相应的测试结果进行调整，其测试的轮次取决于量表度量性能的满意程度。

6. 验证概念性框架

预调查和正式调查都是验证概念性框架的过程。评价概念性框架的适用性主要基于其度量性能，包括信度和效度。

(1) 信度：信度是指在相似条件下所获得的测量结果的一致性，用于评价测量工具的可靠性。PRO 量表常用的信度指标有重测信度、内部一致性信度和测试者内信度。重测信度用于评价量表的可重复性，初测和再测之间的相关系数不宜太低。内部一致性信度用于评价量表的内在一致性，常用 Cronbach's α 系数评价（通常不低于 0.7 为宜）。测试者内信度通常用组内相关系数 (intraclass correlation coefficient, ICC) 评价，有文献报道认为 ICC 一致性可划分成 <0.4 为差，0.4-0.75 为尚可，>0.75 为很好。

(2) 效度：效度是指测量在多大程度上反映了想要测量的内容，用于评价测量工具的有效性。一个好的量表应该既可靠又有效。信度高并不代表效度也高（例如重度抑郁症症状量表用于测量重度抑郁症有较高的信度和效度，而用于测量躁狂症时可能信度高但效度低），但信度低的话，效度必然不会高。

量表效度的评价方法很多，以 3C 方法较为常用，即内容效度 (content validity)、标准效度 (criterion validity) 和结构效度 (construct validity)。内容效度主要基于专家知识和经验以及患者主观判断量表的维度和条目的内容是否合理，是否能正确反映想要测量的内容。标准效度表示研发量表与所谓“金标准”量表的相关程度。由于金标准通常不存在，

且如果存在则研发意义有限（仅在研发量表具有极大的便利性等情况下），因此应用较少。结构效度常通过探索性和验证性因子分析方法评估观测数据产生的结构与概念性框架的一致性。

除了上述 3C 概念外，效度的另一个重要指标是检测变化的能力，又称反应度，即能够灵敏地反映患者结局变化（如干预前后的变化，给予不同干预的反应等）的能力。

7. 撰写量表说明书

为了确保量表的正确使用，应撰写量表使用说明书。量表说明书包括但不限于：目标人群，含引导语在内的完整量表结构，维度和条目的赋值以及量表的计分规则，度量性能，有效应答的规定，缺失数据的处理，回忆期限（如涉及）等。

（二） 用于患者报告结局测量量表的翻译和/或文化调适

临床研究中用于 PRO 测量的原研量表如果为外语，通常需要翻译成中文后才能应用。原研量表的某个或某几个条目如果因为文化方面的差异而无法被患者理解或难以获得有效配合时，还会涉及文化调适问题。量表的翻译和/或文化调适是否恰当，要以翻译和/或文化调适后的量表与原研量表的度量性能是否相近为衡量标准。量表的翻译和/或文化调试可按以下步骤进行：

1.准备阶段。查阅量表研发的所有相关资料；组建多学科翻译团队（如英译中、中译英、医学等专业人员）；建立与量表研发者的沟通渠道，除获得使用该量表最新版本的授权许可外，通过交流更好地理解量表的含义，以使翻译更为准确。

2.正向翻译。两个或多个翻译人员独立将原语言版本的量表译成中文版本，然后综合各个翻译文稿形成中文初稿。

3.回译。由母语为原语言且又熟悉中文的翻译人员将中文初稿翻译回原语言，将回译版本与原文进行比较，如有较大差异，需进一步修改中文译稿，直至回译稿与原文的差异达到可接受的程度，形成中文版初版。

4.量表的文化调适。如果量表中有个别条目不适于当地文化，需对其进行调适，调适结果是否满意应以调适后量表与原研版的度量性能相近为判断原则。

5.中文版初版测试。在目标人群中使用中文版对患者进行认知访谈，评价量表条目的可理解性以及患者的认知程度等，并进行量表性能的定量测试，如果量表的度量性能与原研版相近，中文版可定稿；如果相差较大，则需要进一步完善中文版，直至度量性能达到要求为止，形成中文版终版。

6. 中文版研发报告。中文版终版形成后，撰写完成研发报告，记录整个研发过程，报告度量性能，编写量表说明书，必要时申报中文版软件著作权。

（三） 患者报告结局测量量表的改进

当已有量表并不完全适合研究项目时，应改进后使用。例如，经早期临床试验（如 II 期）数据分析，所用量表不满足研究所需的信度和/或效度，需对量表进行改进或研发新的量表。在开展 III 期试验前应试验对量表再次进行测试，以确保 III 期试验所用量表具有足够的信度和效度。

四、 患者报告结局测量量表的选择与评价

量表作为 PRO 测量工具应具有良好的度量性能，应既可靠又有效。正确选择适用于拟开展的研究项目的用于 PRO 测量的量表甚为关键，结合科学性和可操作性，建议重点关注以下要点：

1. 量表的适用性：考察量表的构建，关注其整体概念是否满足量表研发的目的和符合适用人群，研究目标人群应与原研量表的适用人群一致。

2. 规范文件或系统：是否有规范的量表相关文件或系统，包括但不限于说明文件（特别是量表得分的解释）、用户使用手册、数据收集的标准格式、重要的参考数据（用于设计时的样本量估计）等。

3.研发过程: 量表的使用目的是否明确定义, 研发过程是否严格规范, 量表的结构(维度和条目及其赋权)是否合理, 发表的结果是否详尽。

4.权威性: 研发成果是否在同行评议期刊公开发布, 是否得到较广泛的引用和应用, 是否被指南推荐。

5.语言和文化: 量表的有效性验证是否考虑了不同的教育、文化和种族背景; 新的语种版本是否经过规范的翻译和回译以及验证。经翻译和/或文化调适后的量表其度量性能应与原研量表相近。

6.验证: 是否通过足够大的样本量进行验证, 条目设计和赋值是否合理, 是否有足够的信度和效度。

7.可行性: 量表在使用时的可行性, 包括但不限于实施过程的可操作性、使用多个量表时的条目重叠问题等。患者的应答负担过重可以导致缺失和拒绝应答现象增多, 降低 PRO 数据的质量。增加患者应答负担的因素包括: 量表内容太多, 内容重复性高, 同时选择多个量表且其中某个/些量表意义不大, 量表界面设计不便阅读, 条目涉及不便回答的隐私, 条目设计的不合理等。

五、 临床研究中使用患者报告结局的考虑

(一) 估计目标框架

ICH E9 (R1)中提出的估计目标框架构建的准则和方法对于以 PRO 为试验终点的临床研究同样适用。估计目标框架需在方案和统计分析计划中明确定义。

（二） 选择患者报告结局作为临床研究终点

临床研究如选择患者报告结局作为主要或关键次要终点，应说明选择的理由及依据，结合研究目的、目标适应症的疾病机制、药物作用机理及临床定位等因素综合考虑。对于将 PRO 作为主要或关键次要终点，应注意以下问题：①需要有充分的依据，且与研究目的相一致；②如研究设计未能对患者设盲，会产生较大的主观评价偏倚风险，应极为慎重；③观察期应足够长以体现 PRO 具有临床意义的变化；④应控制整体 I 类错误率；⑤样本量确定应充分考虑预期的差异至少应具有临床意义。

选择的患者报告结局应能反映出患者对药物作用的感受。药物作用不仅限于有效性，也反映在安全性、耐受性或对生存质量的影响等方面，合理选择患者报告结局有助于让研究更好地反映患者体验，使药物研发遵循以患者为中心的理念。

（三） 研究方案和研究报告中有关量表的阐述

使用量表测量的患者报告结局作为主要终点或关键次要终点时，应在研究方案中对其进行说明，包括但不限于：

选择和使用量表的合理性；必要时简要介绍量表的研发和应用情况，特别是针对某些应用较少的量表；量表度量性能的评价方法和指标；量表数据的采集与质量控制；量表数据的分析方法；量表使用的详细说明和培训计划等。

临床研究报告中应包括但不限于：量表数据的收集情况（有效应答、缺失等）；报告所使用量表的度量性能（如信度、效度），并与原研量表比较，当差别比较大时，应分析具体原因和评价对研究结论的潜在影响；量表数据的详尽分析结果以及相应的合理解释。

（四） 量表的有效应答

患者在填报量表时可能会出现缺失、消极应答（如在 5 级 Likert 条目的应答中固定勾选某一级）等现象，从而使得量表的数据失真，因此，量表的使用都应设定有效应答的标准，并在量表使用说明书中规定。例如，某量表规定超过 15%（不同量表有不同定义）的条目未应答，或所有条目都勾选某一级（如“非常满意”）被视为该研究对象的无效应答。在研究方案和/或统计分析计划中需要详细阐明判断有效应答的标准并阐述理由。如果最终判断为无效应答则与无应答一样视为缺失值处理。有些情况下，除了考虑整个量表是否有效应答外，量表的某一维度可能被视为关键变量，此时可能会对维度的应答是否有效事先做出规定。

（五） 缺失数据

PRO 数据特别是量表测量的数据，出现缺失较为常见。因此，研究的实施过程中加强质量控制，尽可能减少缺失十分必要。对于多维度量表中条目数据的缺失，通常会采用填补方法，具体方法优先采用原研量表说明书提供的方法，其次采用文献报道中的主流方法，再次通过当前研究数据的探索性分析确定（通常在探索性研究中完成）。如果不做填补，除了缺失太多被视为无效应答外，需要根据原研量表的规定或事先在方案中定义当条目的分值缺失时处理整个量表和各维度的分值的规则。应在试验设计阶段针对缺失数据制定合理的统计分析策略。

（六） 多重性问题

当 PRO 被列为主要终点之一或关键次要终点时，会涉及多重性问题，其一般处理原则参见《药物临床试验多重性问题指导原则（试行）》。申办者需要在临床研究方案和统计分析计划中事先规定针对多重性问题所采用的决策策略和多重性调整方法。PRO 使用的量表通常包括多个维度，如果其中某个或某几个维度具有重要临床意义，并在方案中被列为关键次要指标（申办者拟在说明书中声称该特定获益），亦会涉及多重性问题，设计时需考虑整体 I 类错误率的控制。

由于量表的多维度和多条目特性，除了侧重于量表整体得分的分析外，各个维度和条目的分析也是必要的，从广义上讲是涉及多重性问题的，但只要它们未被列为主要终点或关键次要终点、或者不在说明书中声称特定的获益，无需进行多重性调整。

（七） 结果的解释

基于量表的 PRO 的结果解释与其它用来评估治疗获益的终点指标相同，阳性结果需同时具有临床意义和统计意义。

最小临床意义差别 (minimum clinical important difference, MCID) 通常用于界定临床意义的阈值，例如，使用 10 分制视觉模拟疼痛量表测量疼痛程度时，干预前后平均分下降多少才有临床意义，或较基线平均下降分值两组的差值大于多少才有临床意义。在确定 MCID 时，应首选相关指南、专家共识等公认的标准；如果没有公认的标准，则需与监管机构及时沟通交流并达成共识，统计方法可能为其提供一定的依据。

采用统计方法估计 MCID，常用的有基于分布的方法和基于锚定的方法。其中锚定法更可靠并且便于跨不同试验进行比较，其根据患者对临床意义的感受设置一个外部的全局性指标（如无改善、轻微改善、显著改善），然后确定对应的量表分值的变化量。通常，全局性指标（等级变量）与量

表分值变化量的相关系数至少达到 0.3 以上才有意义，有研究认为相关系数 0.3 为低度相关，0.5 为高度相关。估计 MCID 还有其它统计方法，如基于混合线性模型的方法等，可与监管机构进行沟通交流后确定主要方法。

（八） PRO/ePRO 的质量控制

应保证研究实施过程中不同研究中心、患者、观察者数据采集的一致性，从而提升临床研究质量。在方案中至少需要明确但不限于：

- 建立质量控制标准操作规程；
- PRO/ePRO 数据采集的时间点和实施顺序；
- 针对相关人员使用 PRO/ePRO 测量工具的培训和指导，包括判断量表完整性的方法和标准，数据填写、存储、传输的时间和方式等，使之充分理解使用量表的目的、量表说明书中的具体内容以及在量表数据收集过程的质量控制环节；
- PRO/ePRO 的数据管理计划。

另外，使用 PRO/ePRO 的临床研究需要更持续主动地现场监查，保证 PRO/ePRO 数据收集的完整性和准确性。

（九） 真实世界研究中 PRO/ePRO 的使用

真实世界研究中，PRO/ePRO 的使用多用于前瞻性研究，如前瞻性观察性研究或实效临床试验。采集的 PRO/ePRO 数

据管理或治理的具体方法参见《用于产生真实世界证据的真实世界数据指导原则（试行）》。

六、 电子化患者报告结局

（一） ePRO 测量

与纸质 PRO 相比，ePRO 在数据收集的高效性、实时性、灵活性、依从性、安全性和患者隐私保护等方面具有明显优势。ePRO 的不足主要体现在某些患者可能会在操作电子设备方面遇到困难，特别是年老、年幼、以及因疾病限制了动手操作能力的患者人群。

目前 ePRO 数据的采集大致有基于电话的交互式语音应答系统和基于屏幕的报告系统两种类型。基于电话的交互式语音应答系统以自动呼叫为特色，运用预先录制的问题以及回答选项脚本，并允许患者使用按键记录应答，数据直接存储到中央数据库。基于屏幕的报告系统可安装在患者自己的电子设备上，如智能手机、平板电脑、计算机，甚至是可穿戴医疗设备，又称为自带设备，病人可访问设备上的网站或软件，根据自身情况选择答案并被记录保存。

ePRO 系统可与电子病历系统或电子数据采集系统对接，形成个体水平的完整数据流；其时间记录功能可有效防止和识别应答回填或提前应答等影响数据可靠性的行为；其远程监控功能有助于研究者、数据管理人员实时进行在线数据管

理和远程数据监测，对有疑问的数据进行质疑标注，及时对受试者进行回访。

（二） 使用 ePRO 的一般考虑

以药物注册为目的的临床研究中，ePRO测量工具及数据采集和数据管理等，应遵循药物临床试验数据管理、电子数据采集、真实世界数据治理相关指导原则的基本要求。

基于网络化平台的ePRO测量方式不同于纸质PRO测量工具，其数据通常上传至在线数据收集中心，供用户综合管理，实现数据存储、监查和导出。因此，为确保研究者具备电子源数据维护和保存权限，研究机构有原始文档支持，以供申办者稽查及监管部门核查，使用ePRO测量工具应遵循以下原则：

1. 研究者应具有维护和确认 ePRO 源数据准确性、真实性的权限。研究者通过稽查轨迹捕捉 ePRO 数据通过测量设备上传后任何数据的变化和修改，避免申办者或第三方机构独自控制原始 ePRO 数据的采集/管理系统。ePRO 源数据是指源于 ePRO 系统最初记录并存储于数据库的记录，如果 ePRO 系统最初记录直接导入 EDC 系统并存储于 eCRF，则最初的 eCRF 为源数据。

2. 数据安全管理体系和访问控制机制。采用加密技术保证数据在收集、提取、传输和存储过程中的完整性、保密性、

可追溯性，防止任何个人或机构修改原始数据，删除患者报告的不良事件、高危预警等数据，建立相对应的访问控制机制，避免计划外揭盲风险。

3. 数据备份。避免试验过程中存在数据损坏或丢失、无法对源数据进行重建或验证的风险。

4. 数据保存。研究机构和研究者应保存有电子源数据或电子文档，使监管部门核查人员可以在临床研究现场检查、核实和复制原始数据。

如果经对研究数据的分析发现ePRO量表度量性能与原研量表有较大差距，应考虑ePRO量表在实施中存在的潜在问题，并予以纠正。此外，基于项目反应理论的ePRO测量工具，通过计算机自适应测试技术，根据前项条目的答案选择后项条目，从而减少条目数量以降低患者应答负担，但减少条目数量应符合保证量表内容的效度的前提。申办者使用此类ePRO测量工具，需提交概念性框架构建、条目库设计筛选流程、程序构建规则以及结果分析解读等相关资料。

七、与监管机构的沟通交流

当申办者计划采用PRO/ePRO作为确证性研究主要或关键次要终点时，应与监管机构及时沟通。沟通的问题包括但不限于目标适应症疾病背景、选择PRO作为主要或关键次要研究终点的理由及依据、研究设计类型、研发量表（如有）

的验证性概念框架及量表说明书等资料、PRO/ePRO改进和/或文化调适（如有）及依据、信度与效度的验证、最小临床意义差别及依据、实施方面的质量控制等问题。进行沟通前，申办者应该向监管机构预先提供包含PRO/ePRO统计分析考虑的试验方案和PRO/ePRO的相关资料。在试验过程中，如果因为更改PRO/ePRO而使临床试验方案做出重大调整，应与监管机构及时沟通。

参考文献

- [1] Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health*. 2003; 6(5): 522-531.
- [2] Bukhari M. “PROMs vs. PREMs (Patient-Reported Experience Measures).”; *Patient Reported Outcome Measures in Rheumatic Diseases*. Ed. Miedany YE. London: Springer, 2016; 405-417
- [3] Byrom B, Watson C, Doll H, et al. Selection of and Evidentiary Considerations for Wearable Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations from the ePRO Consortium. *Value Health*. 2018; 21(6): 631-639.
- [4] Calvert M, Blazeby J, Altman DG, et al, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013; 27; 309(8): 814-822.
- [5] Cohen J. A power primer. *Psychological Bulletin* 1992;

112(1): 155–159

- [6] Coons SJ. ePRO systems validation: clearly defining the roles of clinical trial teams and ePRO system providers. *Value Health*. 2013; 16(4): 457-458.
- [7] Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health*. 2009; 12(4): 419-29.
- [8] Copay AG, Subach BR, Glassman SD, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J Off J North Am Spine Soc*. 2007; 7: 541–546.
- [9] Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. *Health Qual Life Outcomes*. 2010; 8: 89.
- [10] EMA. Reflection paper on the regulatory guidance for the use of health relate quality of life (HRQL) measures in the evaluation of medicinal products. 2005.
- [11] EMA. Reflection paper on expectations for electronic source

data. 2010.

- [12] EMA. Reflection paper on the use of patient reported outcome (PRO) measures in oncology studies. 2014.
- [13] Fayers PM, Machin D. Quality of Life: The assessment, analysis and reporting of patient-reported outcomes (3rd Edit). John Wiley & Sons, Ltd. 2016.
- [14] Ferreira ML, Herbert RD, Ferreira PH, et al. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. J Clin Epidemiol. 2012; 65: 253–261.
- [15] FDA. Clinical outcome assessment (COA) compendium. 2021.
- [16] FDA. Clinical outcome assessment (COA) qualification program. <https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program>.
- [17] FDA. Guidance for industry: Assessing COVID-19-Related Symptoms in Outpatient Adult and Adolescent Subjects in Clinical Trials of Drugs and Biological Products for COVID-19 Prevention or Treatment. 2020
- [18] FDA. Guidance for industry: Electronic source data in

clinical investigations. 2013.

- [19] FDA. Guidance for industry: Patient-Reported Outcome Measures: use in medical product development to support labeling claims. 2009.
- [20] FDA. Patient-Focused Drug Development: collecting comprehensive and representative input. 2020.
- [21] FDA. Roadmap to patient-focused outcome measurement in clinical trials. 2015. <https://www.fda.gov/media/87004/download>
- [22] FDA. Plan for issuance of patient-focused drug development guidance. 2017.
- [23] FDA. Upper facial lines: developing botulinum toxin drug products. 2014.
- [24] Fiero MH, Pe M, Weinstock C, et al. Demystifying the estimand framework: a case study using patient-reported outcomes in oncology. *Lancet Oncol* 2020; 21: e488–94
- [25] Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*, 1975 ; 31: 651-659.
- [26] Fox MW, Onofrio BM, Onofrio BM, et al. Clinical outcomes and radiological instability following decompressive lumbar laminectomy for degenerative spinal stenosis: a comparison

- of patients undergoing concomitant arthrodesis versus decompression alone. *J Neurosurg*. 1996; 85(5):793-802.
- [27] Hong K, Majercak KR, Villalonga-Olives E, et al. Patient-reported outcomes in breast cancer FDA drug labels and review documents. *J Patient Rep Outcomes*. 2021; 5(1):36.
- [28] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415.
- [29] Lawrance R, Degtyarev E, Griffiths P, et al. What is an estimand& how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials? *J Patient Rep Outcomes*. 2020; 4(1):68.
- [30] Ly JJ, Crescioni M, Eremenco S, et al. Training on the use of technology to collect patient reported outcome data electronically in clinical trials: best practice recommendations from the ePRO Consortium. *Ther Innov Regul Sci*. 2019; 53(4): 431-440.
- [31] Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement Instruments. *BMC Med Res Methodol*. 2006; 6: 2.

- [32] Walters S. Quality of life outcomes in clinical trials and health-care evaluation: A practical guide to analysis and interpretation. John Wiley & Sons, Ltd. 2009.
- [33] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. Value Health. 2005; 8(2): 94-104.
- [34] Guideline, ICH. "Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials." E9 (R1). Step 4 (2019): 20.

附录1：词汇表

标准效度 (Criterion Validity): 又称校标效度, 指研发的PRO量表的评分与已知所谓“金标准”量表对同一概念的度量的相关程度。大多数PRO量表因无金标准而无法衡量其标准效度。

测量工具 (Instrument): 一种获取数据以及支持其使用的所有信息和文档的工具(如量表), 通常包括实施方面的详尽指引、数据收集的标准格式、用以说明计分和分析方法以及目标疾病人群的结果解释的规范文件等。

概念 (Concept): 又称感兴趣的概念(concept of interest, COI)。在监管层面, 概念是PRO量表所捕捉或反映的个体在临床、生物、生理、功能等方面的状态或体验。在PRO层面, 概念代表了患者对其健康状况或与治疗相关的功能或感觉。

患者报告结局 (Patient-reported Outcome, PRO): 为任何来自患者直接报告且不被他人修改或解读的对自身疾病和相应治疗感受的评估结局。

患者为中心的药物治疗(Patient-focused Drug Development, PFDD):指一套系统的方法, 在药物全生命周期中, 该法有助于确保患者的体验、观点、需求和优先顺序能够被获取并有效地融入到药物的研发和评估中。

回忆期限 (Recall Period): 患者在应答当时距所要回答的PRO 条目或问题的时间。回忆可以是瞬间的(实时的), 或

前一段时间的追溯。回忆期不宜太长，如一般不超过一周。

检测变化的能力(Ability to Detect Change): 测量工具检测PRO测量得分随测量条件(干预前后不同时间点、不同干预、不同人群等) 变化而呈现差异的能力。

结构效度 (Construct Validity): 又称构建效度，指由观测数据所呈现的PRO量表的条目、维度和所要表达的概念之间的结构关系是否与量表研发的理论构想相符。

Cronbach's α 系数: 用于评价量表内在一致性的信度指标。

量表的概念性框架 (Conceptual Framework of a Scale): 基于既往研究（文献）、专家知识和经验以及必要的调研所构建的量表的维度和条目的框架。维度的数量和命名依据对研究内容的理解设定，每个维度下的条目数和条目内容用以体现其所属维度的内涵和重要程度（例如，每个条目等权时，维度下的条目数量就体现了维度的重要性）。

内容效度 (Content Validity): 基于专家知识的定性研究，验证量表是否能够测定其所希望测定的内容。

生存质量 (Quality of Life, QoL): 又称生命质量，生活质量，用于评估生活各个方面所体现的整体健康状况。

条目 (Item): 某一问题、陈述或任务（以及标准化的应答选项），被用于患者对特定概念的评估。

调适 (Adaptation): 基于种族间语言和文化差异的考虑对量表进行的任何更改。调适不会改变PRO量表的结构，但会调整小部分内容以适用于另一种模式、语言或人群。调适研究是为了验证PRO量表在新环境或新语言下的度量性能。

维度 (Domains/Dimensions/Factors): 构成量表的第一级结构（二级结构量表）或第一和第二级结构（三级结构量表），用于表达组成量表的某一方面内容（概念）。一个维度由一个或多个条目构成。

效度 (Validity): 指测量在多大程度上反映了想要测量的内容，用于评价PRO量表的有效性。

信度 (Reliability): 指在相似条件下所获得的测量结果的一致性和可重复性，用于评价PRO量表的可靠性。

症状 (Symptom): 只能由患者察觉和感知的疾病、健康状况或治疗效果的任何主观证据。

治疗获益 (Treatment Benefit): 治疗对患者生存、感觉或功能的影响。治疗获益可以通过有效性或安全优势来证明。例如，治疗效果可通过症状进展方面的改善或延缓来测量，也可通过减少或延缓治疗相关毒性来测量。不能直接获取治疗对患者生存、感觉或功能的效果的测量是治疗获益的替代测量。

最小临床意义差别 (Minimum Clinical Important Difference,

MCID): 通常用于界定临床意义的阈值，例如，使用10分制视觉模拟疼痛量表（VAS）测量疼痛程度时，干预前后平均分下降多少才有临床意义，或较基线平均下降分值两组的差值大于多少才有临床意义。

附录2：中英文词汇对照

中文	英文
标准效度	Criterion Validity
测量工具	Instrument
测试者内信度	Intra-rater Reliability
重测信度	Test-retest Reliability
电子化患者报告结局	Electronic Patient-reported Outcome (ePRO)
电子数据采集	Electronic Data Capture (EDC)
度量性能	Measurement Properties
概念性框架	Conceptual Framework
患者报告结局	Patient-reported Outcome (PRO)
患者为中心的药物治疗	Patient-focused Drug Development (PFDD)
回忆期限	Recall Period
量表	Scale
简化版口腔干燥量表	Summated Xerostomia Inventory, (SXI)
交互式语音应答系统	Interactive Voice Response Systems (IVRS)
基于屏幕的报告系统	Screen-based Reporting Devices
基于锚定的方法	Anchor-based Method
检测变化的能力	Ability to Detect Change

中文	英文
结构效度	Construct Validity
内部一致性信度	Internal Consistency Reliability
内容效度	Content Validity
生存质量	Quality of Life
视觉模拟标度	Visual Analog Scale (VAS)
条目	Items
条目池	Item Pool
调试	Adaptation
维度	Domains/Dimensions/Factors
项目反应理论	Item Response Theory (IRT)
效度	Validity
信度	Reliability
应答负担	Respondent Burden
组内相关系数	Intraclass Correlation Coefficient (ICC)
症状	Symptom
重度抑郁症症状量表	Symptoms of Major Depressive Disorder Scale (SMDDS)
治疗获益	Treatment Benefit
自带设备	Bring-Your-Own-Device (BYOD)
最小临床意义差别	Minimum Clinical Important Difference (MCID)