欧盟制药工业近红外光谱技术应用、申报和变更资料要求指南(草案)*

□乔延江** 杜 敏 史新元 吴志生 徐 冰

(北京中医药大学中药学院 北京 100102)

摘 要:本文旨在介绍欧盟制药工业中,近红外光谱法用于定性、定量分析以及过程监控时方法校正、验证和维护的指导原则,以及申报与变更时所需呈递的资料要求。以期为我国近红外光谱技术在制药领域中的研究与应用提供参考。

关键词:近红外光谱(NIRS) 过程分析技术(PAT)

doi: 10.3969/j.issn.1674-3849.2012.04.034

目前,近红外光谱技术(Near Infrared Spectroscopy, NIRS)已被广泛应用于制剂分析、化学分析、物理分析和过程分析等领域,如各种原料药、中间体和成品的定性定量分析。在ICH Q8质量源于设计(Quality by Design, QbD)原则的指导下,近红外光谱技术作为过程分析技术(Process Analytical Technology, PAT)之一,通常用于制剂生产过程的实时放行测试(Real Time Release Testing strategy, RTRT)。在实际应用中,近红外光谱分析方法的建立与很多因素有关,如测量原理、化学计量学方法、仪器类型等。同时,近红外光谱分析方法的开发与应用并非一劳永逸,而是不断发展和完善的过程,且存在自身的生命周期。

为进一步推广并规范近红外光谱技术在制药领域的应用,欧洲医药管理局颁布了《欧盟制药工业近红外光谱技术应用、申报和变更资料要求指南(草案)》(以下简称《指南》),围绕近红外光谱分析方法的建立、校正、验证、维护和应用等,提出了一系列指导原则和要求。本文对此予以详细介绍,以期为我国近红外光谱技术在制药领域中的研究与应用提供参考。

一、概 述

为方便描述,《指南》中采用以下术语:近红外光谱采集方法(NIRS method):主要描述影响测定待测组分的近红外仪器方面的因素,包括样品测定方式和仪器参数设置等;近红外分析模型(NIRS model):描述近红外光谱与待测组分之间的相关关系,一般采用化学计量学方法来建立;近红外光谱分析方法(NIRS procedure):描述在界定范围内,如何应用近红外光谱采集方法和近红外分析模型。

近红外光谱法无需或很少对样本进行预处理,可实现高效、快速的测定。但近红外光谱数据非常复杂,必须借助化学计量学模型才能对其所包含的

收稿日期:2012-07-09

修回日期:2012-08-13

* 科学技术部国家"十一五"重大新药创制专项(2010ZX09502-002):清开灵注射液安全性关键技术研究,负责人:乔延江。

** 通讯作者:乔延江,本刊编委,教授,博士生导师,北京中医药大学副校长,主要研究方向:中药信息工程,中药新药研发及中药生产过程在线监测,Tel:010-84738621,E-mail: yjqiao@263.net。

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

信息进行解读。在建立化学计量学模型时，需选取代表性样本，并采用标准样品和参考分析方法对样品进行分析，然后建立参考值和近红外光谱之间的关系模型作为预测的基础。这也是近红外光谱法不作为基础分析方法使用的原因。

在目前的原料药和药品的放行测试中，仍以常规分析方法为主，近红外光谱法用于放行测试时需要与参考值测定方法相互配合。在某些工艺（如混合）的过程分析中，常规分析方法存在较大的测量误差，则需优先考虑采用近红外光谱分析技术。

近红外光谱分析方法的建立与完善是一个动态过程，而且过程的各个阶段是相联系的（如图1）。对于近红外光谱分析方法而言，有些阶段如“方法变更”或“PAT 应用阶段”并非是必需的，但需对缺省的阶段进行说明。当购得或生产出新批次的样本时，可选择新样本加入到近红外分析模型中对模型进行更新。有关变更实施的指导准则，参见第 7 部分。

二、适用范围

《指南》内容涉及使用近红外光谱技术的人用或兽用药品上市许可申请和变更申请的法规要求。近红外光谱法常需借助于化学计量学方法对待分析物的响应信号进行解释，因此有别于 HPLC 或 GC 等常规分析方法。近红外光谱法在《欧洲药典》中有相应描述（Ph. Eur. 2.2.40），但仅以此不足以作为药品上市许可申请或变更申请时使用此方法的支撑。

《指南》同时概述了近红外光谱法应用时的一般要求，包括定性、定量过程，或作为过程分析技术用于监控药物合成和药品生产的过程。如果进行适当的解释或具有合理性，也可采用《指南》以外的其它方法。《指南》中阐述的化学计量学原理也适用于其它分析技术。

三、法律依据

阅读《指南》时可参阅欧洲议会和理事会指令 2001/82/EC（修订案）和指令 2001/83/EC（修订案），以及以下各项：

·欧洲药典 2.2.40

·ICH Q2 (R1) 分析方法验证指南(CPMP/ICH/381/95)

·VICH 分析方法验证指南 GL1 与 GL2 (CVMP/VICH/590/98 与 CVMP/VICH/591/98)

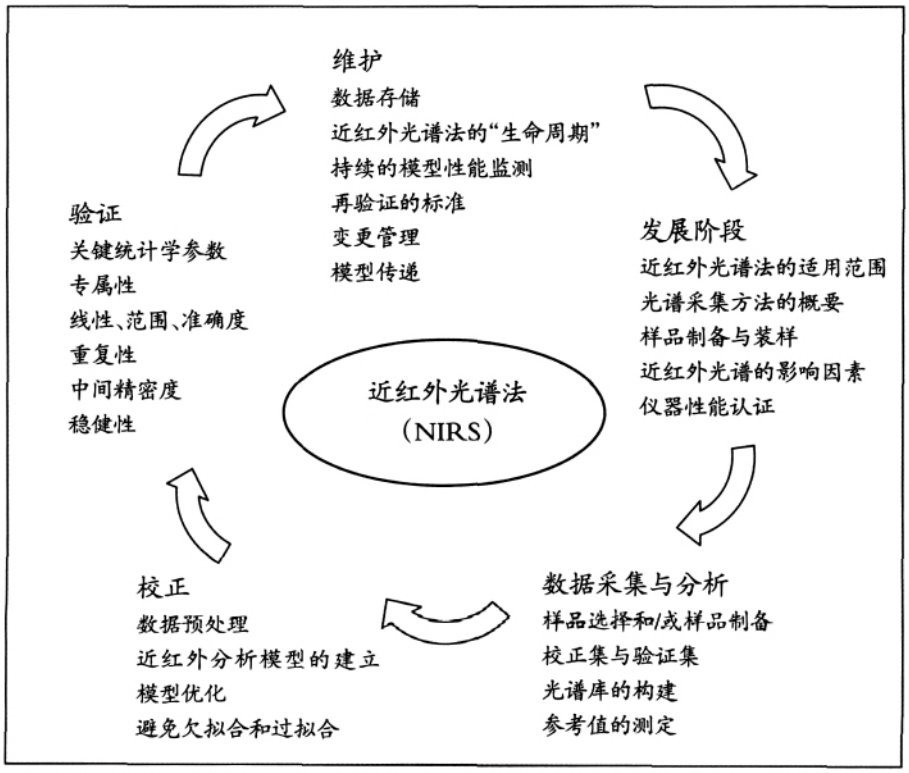

图 1　近红外光谱法的建立流程

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

·专利药物委员会 (CHMP) 与畜用药物委员会 (CVMP) 有关工艺验证指南的意见(CPMP/QWP/848/96 与 EMEA/CVMP/598/99)

·ICH Q8: 药品研发指南

·ICH Q9:质量风险管理指南

·ICH Q10:药品质量体系指南

·ICH 对 Q8、Q9 和 Q10 有关问题的解答(CHMP/ICH/265145/2009)

四、一般要求

1. 概 述

(1) 近红外光谱分析范围的确定。

采用近红外光谱法时,应界定清楚近红外光谱分析方法的应用范围(即对于既定目标采用哪种近红外光谱采集方法和近红外分析模型),这对后续方法变更尤为重要。近红外光谱分析方法的范围涉及近红外光谱采集方法中的细节部分(即仪器、样品测定方式等关键因素)、近红外分析模型(即近红外光谱与待测组分或待测性质的关联关系)以及方法应用的限制性因素(如分析浓度、探头位置、波段范围、化学计量学算法等)。考虑到未来如何从监管的角度进行方法的调整,对近红外光谱分析方法应用范围的界定至关重要(参见第 7 部分)。

近红外光谱广泛应用于定性和定量分析工作中,应用时需清楚地理解光谱测量的物理化学基础、仪器及化学计量学方法,并了解方法建立时做的所有假设。

在测量待测组分时,近红外光谱信号有可能受到散射光的影响。因此,申请人必须对近红外光谱方法的应用范围和目的进行讨论,并证明其与待测组分或待测性质的相关性。

近红外光谱法的一个应用前提是:对适用范围之外的样品(如浓度超出范围,组成不同等)能够进行正确识别,以避免采用所建方法得到错误结果。

近红外光谱所含的信息非常复杂,常需借助于化学计量学方法从光谱中提取相关信息或证明光谱值与化学参考值间的关联关系。化学计量学数据分析和建模通常采用统计软件实现。通常,将光谱信号中的变异与潜变量进行关联时会受到校正参考值的制约。通过软件建立的相关关系可能只是一个偶然而非待测组分浓度变化的结果,因此这些校正模型常需采用独立样本集进行验证。

(2) 近红外光谱采集方法概述。

申请人应提供近红外光谱采集方法的详细内容,包括光谱采集模式(如反射、透射、透反射)、单色器的类型(如光栅、FT-IR)、波长(波数)范围、检测器类型(如硅、硫化铅)、光学带宽、光谱分辨率、波长准确度和精密度、信噪比、取样装置以及分析过程中其它必要的部件和控制方法。此外,还需说明数据采集、分析和相关统计软件的使用方法。

(3) 可行性研究。

在建立一个新的分析方法时,需对近红外光谱法的可行性进行研究,以考察其是否适于分析目的。可行性研究至少应包括考察合适的光谱响应值、方法专属性、样本基质的干扰等,并考察样本处理与制备的影响。

(4) 影响光谱响应的因素。

在进行近红外光谱采集时,可能存在影响光谱响应的背景理化因素。将其全部列出是不可能的,影响因素主要包括测量环境、样本温度、残留水分和溶剂、样本厚度、样本的光学特性、样本杯的光学性质、样本晶型、粒度、样本均匀性和样本陈化程度。同时还需考虑测量时间和仪器的基线漂移。

申请人需对所有影响光谱响应的潜在因素依次进行讨论,并证明各因素的影响不显著或已得到满意的控制(需提供支持数据)。

2. 数据采集

(1) 样本制备与装样条件。

如果样本需要制备,则需提供该制备过程的详细信息,并证明其合理性;应描述样本与近红外光谱检测器间的接触界面;详细讨论装样条件对近红外光谱的影响,并附以数据支持。如果装样条件对近红外光谱具有显著影响,应证明可对其进行有效控制。

在进行近红外光谱采集前,优化样品装样条件非常重要,主要包括对样品形态、粒径、样品杯的光学性质以及环境条件的优化。还需对所用光谱范围及样本累积扫描次数进行说明并证明其合理性。

(2) 样本总体。

用于近红外光谱分析的样本按照可行的方式进行取样,能代表生产过程,也可采用在研发和中试过程中获取的能够代表市售产品的样本,但应给出能够支持样本选择的合理依据。

对于定性或定量分析方法,其样本总体应涵盖

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

常规生产中可能遇到的所有潜在变异。这些变异包括：待测组分的浓度、粒度、原材料来源、水含量、残留溶剂、样本基质中的定性和/或定量变异因素（如辅料、处方等）、过程变异、样本陈化程度、温度等。

(3) 数据预处理。

由于近红外光谱受粒度、装样条件等物理因素的影响，在模型构建和验证之前需要对原始光谱进行预处理以减少噪音并强化谱带特征。常用数据预处理方法包括归一化法和导数法等。

进行数据预处理时须要谨慎，因为预处理可能会导致有效信息的损失。因此对任何数据进行预处理都应作好记录并进行解释说明。

(4) 参考分析方法（可选）。

建立近红外光谱法时，常需采用参考分析方法对校正集样本和验证集样本进行鉴定或化学值测定。理想情况下，参考值的测定应与近红外光谱采集同时进行。

选用参考分析方法时，需提供相应的支撑数据，具体包括：对分析方法的描述，需依据 ANDA 申报材料 3.2.P.5.21 的数据要求；参考分析方法的详细验证信息，需依据 ANDA 申报材料 3.2.P.5.31 的数据要求和 ICH Q2（R1）分析方法验证指南（CPMP/ICH/381/95），对于兽药，需依据 VICH 分析方法验证指南 GL1 & GL2（CVMP/VICH/590/98 & CVMP/VICH/591/98）；相关参考标准和物质的详细信息，需依据 ANDA 申报材料 3.2.P.61 的数据要求。

(5) 构建参考光谱库。

参考光谱库的组成应涵盖近红外光谱分析方法的适用范围，并符合变更管理制度（符合 GMP 验收标准）。光谱库中的样本批次应能代表市售原料或产品，并标明批号。

对于定性分析，其参考光谱库可能非常庞大且类型多样，因此有必要将光谱库划分成适当的子光谱库，以避免校正模型太过复杂。对子数据库的选择和子数据库数目的确定，应予以描述和解释。

3. 数据解析

(1) 近红外分析模型的描述。

对所采用的近红外化学计量学模型应予以充分描述。关于校正模型的更多信息见“定性分析方法”和“定量分析方法”部分。

(2) 光谱质量统计检验。

采用近红外分析模型对未知样本进行检测时，首先应对样本光谱的质量进行统计学评价（即模型适用性检验），以确定待测样本是否处于校正模型的适用范围内。通过统计学评价，如 Hotellings T^2 检验或 DModX（Distance to Model）图，可以考察样本光谱是否处于预先设定的变异范围内或样本是否为离群值。对离群值的检出方法应进行描述（见一般要求中“离群值”）。若样本为离群值，采用所建校正模型进行样本的检测可能得到“假阳性”或不正确的结果，因而不具有科学性。

采用任一近红外分析方法时，均需进行光谱质量检验，该检验通常由计算机软件自动执行。

(3) 离群值。

对于样本数据中的可疑离群值（近红外光谱值或参考值），无论是将其纳入校正集、测试集或验证集还是从中剔除，都需进行研究并提供依据。

《指南》中“离群值”是指非预期结果或不在特定范围内的结果。它可能是“光谱离群值”（即光谱数据为离群值但预测结果在特定范围内），或“参考离群值”（即光谱值在特定范围内但化学参考值为离群值），亦或光谱数据和参考值均为离群值。

导致离群值产生的原因很多，如：待测样本与建模样本不属于同一样本总体，仪器故障，参考分析方法出错或者数据记录错误等。从这个意义上讲，离群值可能不一定是测定错误，而仅仅是测定结果与其它样本不同，而这种不同可能会对分析结果产生不确定的影响。

如果待测样本由于自身的独特性质而被判定为离群值，则需采用适当方法对其进行验证。经验证后可将该样本加入光谱库，并对模型进行重新验证，以此添加原模型中不包含的新信息。这是近红外光谱分析方法“生命周期”（见图 1）的关键部分，该过程可确保近红外光谱法得到及时更新和优化。

在近红外光谱分析方法建立或再建立阶段、校正或验证阶段，都需采用恰当的方法处理数据中的离群值，并对这些方法进行阐述。

(4) 常规分析中的不合格结果。

采用近红外光谱法进行常规分析时，如出现不合格的结果，则需按照本公司的药品质量控制规范对不合格样本进行考察。对于近红外分析不合格，而参考分析方法结果合格的样品也不能放行。

如果经检查后，不合格样本处于参考方法的控制范围内，那么说明所用的近红外方法校正不充

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

分。因此需要对近红外光谱法进行更新(对近红外光谱法整个"生命周期"进行更新,见图 1),并重新对不合格产品进行分析,以实现样本放行。

4. 校正过程

对校正过程的具体要求见 "定性分析方法"和"定量分析方法"部分。

5. 验证过程

对近红外光谱分析方法的验证过程需符合 ANDA 申报材料 3.2.P.5.32 中的数据要求和 ICH Q2(R1) 分析方法验证指南(CPMP/ICH/381/95)。对于兽药,需依据 VICH 分析方法验证指南 GL1 & GL2(CVMP/VICH/590/98 & CVMP/VICH/591/98)。

验证集样本(外部验证)应与校正集完全独立(见 5.5 和 6.4)。验证近红外光谱法时需将近红外分析结果与参考分析方法的结果进行比较,并结合中间精密度等其它参数。总之,无论是否采用参考方法,都应明确近红外方法验证的可接受标准,并根据应用目的作针对性的调整。

在近红外光谱法的初始注册文件中,如采用了参考分析方法,应包含该方法的验证数据(见一般要求中"参考分析方法")。如果近红外光谱法只是作为上市许可产品被批准的参考分析方法的补充,则需提供符合 (V)ICH 分析方法验证准则的参考方法验证数据。有关验证过程的具体要求见分章节"定性分析方法"和"定量分析方法"。

6. 近红外光谱法在过程分析技术中的应用

近红外光谱分析方法用于过程分析时大都专属于特定的生产单元。因此在《指南》中并未对此类近红外光谱分析方法规定具体要求。《指南》中对数据的一般要求同样适用于 NIRS 过程分析,但应兼顾 NIRS 过程分析方法的预期目的和应用范围。

对 NIRS 过程分析时的光谱采集方法需进行详细描述,如光谱仪的空间放置位置、生产设备的结构特点以及取样方式等。

NIRS 过程分析方法所需的信息量取决于其预期目的。如缺少对过程分析方法的描述或缺少对过程的控制 (如取样方式、NIRS 模型、校正过程和验证过程等),则应对其进行科学的描述与解释。

NIRS 在化学药品生产过程中的应用包括:原料药生产过程如化学反应动力学、结晶、干燥和研磨过程;药品生产过程如制粒、混合、压片和包衣过程。

在监测粉末混合均匀度时,可通过测量 NIR 光谱值随时间的变化(如光谱的标准偏差)来监测混合过程,并不需要采用 HPLC 等参考方法。

7. 资料的一般要求

以下资料要求适用于 NIRS 定性和定量分析方法(有关定性分析和定量分析的具体要求分别见第 5 和第 6 部分):近红外光谱分析方法的范围(见一般要求中"近红外光谱分析范围的确定");校正集、测试集及验证集样本的具体组成;对参考分析方法的描述(若使用参考分析方法);近红外光谱分析方法和参考方法的校正和验证过程的报告(如果采用了参考分析方法且已被批准,给出具体内容的概述);对近红外光谱分析方法生命周期进行维护的方法。

如果申请人希望使用《指南》中未包含的术语,需对新术语进行充分、清晰的解释说明(如使用词汇表的形式)。

五、定性分析方法

1. 概 述

近红外光谱法广泛应用于定性分析,其应用大致可分为 3 个方面,即鉴定、判别和合格检查。

(1) 鉴定与判别。

在《欧洲药典》中,鉴定(Identification)是指确认某种化学物质的存在。在制药工业中,鉴定的概念比较广泛,有时指对同一化学物质不同物理属性(如粒度、晶型等)的区分。为方便区别,《指南》中"鉴定"仅指对化学结构差异的区分,"判别"指对化学与物理属性差异的区分。

采用近红外光谱法对某一物质 (如原料药、辅料、混合物、药品、中间体等)进行鉴定或判别,是基于待测样本的光谱与光谱数据库中多个批次不同样本的光谱之间的比较。为实现正确判别(通过,不匹配或结果不确定),需采用化学计量学方法并设定合理的置信水平。

如果所得结果模棱两可,应修正近红外方法或在分析方法应用范围内去除干扰物质以保证对待测物质的正确判别。

对物质的分类可分步进行。例如,可先对某一化学成分组(Chemical identity)或一组相关物质进行鉴定,然后利用更具选择性的数据库对每一个物质进行判别。这种方法可降低假阳性率和假阴性率。

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

通常在对样本进行鉴定之后，再进行判别。

如果单独采用近红外光谱法不足以对某一组分进行鉴定或判别，应辅以其它分析方法（如化学反应或色谱法），通过多种分析方法的结合尽可能保证方法的专属性。

(2) 合格检查（PAT，动态过程监测或趋势分析）。

《指南》采用“合格”(Conformity)一词表示过程特性（物理和/或化学属性）在一定程度上合乎特定标准。采用近红外光谱法进行的合格检查主要针对能否产生连续变量的生产过程。合格检查是指对过程特性的描述或过程变化趋势的分析，如通过监测NIRS信号的变化以确定过程终点，因此这类方法也被称为“动态过程监测”、趋势分析或过程分析（见“近红外光谱法在过程分析技术中的应用”）。由于这一过程中取样进行参考分析具有较大难度，因此通常不涉及采用参考分析方法。

合格检查与定性分析方法在校正和验证方面很类似，但其校正与验证的具体内容根据方法的预期目的而定。例如，对“移动块标准偏差”法的验证主要针对具体的过程终点，并有正确的理论依据和对该方法预测能力的分析证据作为支撑。

2. 数据采集

(1) 样本采集与样本总体。

样本如何选择以及光谱数据库是否需要后续扩充取决于近红外光谱分析方法的复杂程度。所有样本均需采用现行标准方法或其它适当方法进行检验。

近红外光谱数据库中应包含足以代表未知样本正常变异的基本批次数目。为扩充生产样本的变异性，有时需要在实验室条件或中试规模制备新样本。除有合理的说明外，这些样本必须采用与生产规模样品相同的方法制备。自制样本与生产样本的比例应结合常规生产中的预期变异而定。建模样本的选择应足以确保近红外光谱法在常规分析中的稳健性。对于原材料的鉴定或判别，数据库中应包含所有潜在供货商的样本。对于合格检查（用于过程控制与监测），样本总体应根据分析方法的目的来合理确定。

(2) 样本量。

光谱数据库中的样本量主要取决于样本基质的复杂程度以及待分析组分受基质的干扰程度。通常样本基质越复杂，样本量应越大以覆盖统计总体。

校正集与验证集中的批次数量及每批次的样本量应足够、合理且包括生产过程中的所有变异因素。

(3) 样本集的组成。

在建立、优化和验证近红外定性分析模型时，通常需要两个样本集，即校正集和独立的验证集。前者用于建立校正模型；后者用于对选定模型进行（外部）验证。

校正集样本包含组成光谱数据库的所有样本。最简单的光谱数据库由不同种类的所有样本组成，但其专属性不高。通常可采用包含某一特定等级所有样品的子数据库以满足所需专属性。一般可先利用主数据库对物质进行鉴定，再利用子数据库进行判别（如对一个已经鉴定的物质进行晶型判别）。

校正集与验证集样本在近红外分析方法的适用范围内应具有代表性，且所含样本能覆盖其所在样本总体的整个变异范围。其中验证集应完全独立于校正集，并包含阳性样本与阴性样本。

以上原则同样适用于合格检查过程。其建模样本应覆盖所监测过程中样本的变异范围，并能准确测定过程终点。验证集样本应包含阳性样本及阴性样本，以确保近红外光谱分析方法适用于使用目的（例如，测定混合过程均匀程度时，其样本集中应包含混合不充分的样本及过度混合以致分层的样本）。

(4) 离群值。

为实现有效分析，应检出并剔除光谱数据库中的离群值。对这一过程应记录并证明其合理性。参见“一般要求”中“离群值”。

3. 模型校正

定性校正方法包括主成分分析（Principal Component Analysis, PCA）、判别分析（Discriminant Analysis, DA）、簇类的独立软模式法(Soft Independent Modelling of Class Analogues, SIMCA)、聚类分析（树图）、K-最近邻法（k-Nearest-Neighbourhood-Analysis, KNN）、支持向量机（Supported Vector Machines, SVM）和距离匹配法等相关算法。

校正算法的选择取决于光谱库的应用范围。一般而言，在保证结果精度的前提下，选用尽可能简单的算法。

为对样本进行正确鉴定和/或判别，通常需设定阈值、置信限或容差。在验证报告里，对校正模型相

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

关值的设定应进行解释说明并证明其合理性。

4. 模型优化

通常定性分析方法的优化包括建模样本的选择、数据预处理方法及建模方法的筛选3个方面。

5. 模型验证(内部验证与外部验证)

(1) 概述。

近红外定性分析方法验证的主要内容是方法的专属性和稳健性。验证方式包括内部验证和外部验证。

内部验证是指运用重采样方法进行交叉验证,通过循环抽样得到光谱数据库的子样本集或多个子样本集进行多次交叉验证。内部验证过程应能证明参考光谱库中的所有样本在设定阈值、置信限和/或容差范围内可被正确鉴定或判别。外部验证则是采用未包含在参考光谱数据库中的样本集来验证校正模型的预测性能。

定性分析方法的验证需满足专属性和稳健性的基本要求。

(2) 专属性。

专属性实验的内容依近红外光谱分析方法而定。若近红外光谱分析方法专属性不高,可采用其它分析方法予以支持补充。参比库中光谱所对应的独立样本必须全部经过检验并认证。

应采用易混淆的物质来检验参考光谱库,且这些易混物质应能被正确识别。对药品进行鉴定或判别时,如无特殊理由,验证集中应含有名称类似或结构类似的易混物质。如果可行(如对于定性分析),验证集中应该包含同一物质不同等级的混淆品、无水/水合物质、不同晶型物或来源于不同供货商的物质。

应证明对于每一个检测参数,在检出合格批次与不合格批次时近红外光谱分析方法与参考方法具有相同的选择性。应清晰描述(外部)验证集的组成,并说明其合理性。

(3) 稳健性。

进行近红外光谱分析时,应充分了解影响分析结果的因素,并对其进行检测和记录,包括温度(包括环境和样本的温度)、湿度、样品在光学窗口中的位置、装样装置、样品瓶/样品管的差异、光纤探头的深度,乃至包材的差异。验证模型稳健性时,需考虑到仪器间的差异,如光源或标准物质的更换等。

可考虑采用实验设计(Design of Experiments, DOE)来使有效信息最大化。

六、定量分析方法

1. 概 述

考虑到近红外方法开发商的利益,应对近红外光谱定量技术进行可行性研究以证明其可行性;并概括建模过程及验证过程的基本要求(如测定制剂均匀度时,为确保测定的是总含量而不是浓度,就必须考虑片重这一因素)。

2. 数据采集

(1) 样本采集及样本总体。

如果可行,校正集应既包含生产规模样本,又包含用来模拟样本潜在变异的实验室或中试规模制备的样本。为了按照规范要求正确评价方法的线性,有时需补充实验室自制样本来扩大生产样本的范围,这些自制样本必须按照正常生产流程进行制备。样本集中生产批次样本与自制样本的比例应结合常规生产中的预期变异而定。

为满足回归相关统计的基本假设,避免产生偏差,应确保样本在整个潜在变异范围内均匀分布。同时,应考虑到引入偶然和系统误差的可能性。

对样本分布的评价应结合近红外方法的预期目的。并应对样本及样本数量的选择进行解释,如果可能,需补充与应用目的相关的实验设计(DOE)。

(2) 样本量。

校正算法一般建立在光谱变异与主成分间关联关系的基础上,但会受到校正参考值的制约。为避免偏差,建模样本数应远远大于所用的主成分数(如果可以,两者数目也可相当)。任何情况下,都应对样本数的选择进行解释说明。

建立定量校正模型所需的样本数取决于样本基质的复杂程度和/或待测组分受样本基质的干扰程度。一般情况下,样本基质越复杂,需要越多的样本以覆盖统计总体。如样本基质中仅包含两种简单组分,其所需的样本数就要少于包含多组分的复杂体系。对后者而言,所建模型更为复杂,包含的主成分数也越多,因此需要更多样本以确保其有效性。

校正集与验证集中所包含的批次数量,及每批次包含的样本数应足够合理。

(3) 校正集,测试集及验证集样本组成。

为了建立、优化和验证近红外定量校正模型,

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

通常需要3个样本集(与定性模型相类似,但名称有所不同):校正集,用于建立校正模型;测试集,用于优化和筛选校正模型;验证集,用于对选定模型进行外部验证。

校正集样本用于建立待选校正模型,样本集中应包含能体现近红外光谱分析方法使用范围内所有潜在变异的样本。优化或筛选模型时往往采用内部验证。内部验证是指运用重抽样的统计方法进行交叉验证,通过对模型性能进行内部验证以达到优化的目的。通过循环抽样得到光谱数据库的子样本集或多个子样本集进行多次交叉验证以筛选出适于现有数据的最佳校正模型。

测试集用于对校正模型的有效性进行初步测试或检查。由于测试集与校正集为相同批次总体中不同批次的样本,它并不能对校正模型进行独立验证(独立验证必须采用完全独立的样本集来完成)。实际上,通常将样本总体的2/3用做校正集,剩余1/3作为测试集,校正集中应含有足够多的样本以确保所建模型的稳健性。应对校正集和测试集的组成、数量进行说明,并证明其合理性。

验证集(用于外部验证)是完全独立的第3组样本集,它与校正样本集为不同的批次总体。原则上,外部验证集应涵盖样本总体的整个变异范围,并包含生产规模的样本批次。对验证集中所含的批次数、批次规模和批次组成应进行讨论与解释。

验证集样本用于验证校正模型,并以统计学参数—预测标准差(The Standard Error of Prediction, SEP)来评价校正模型的有效性和预测能力。对校正集也可计算得到校正标准差(The Standard Error of Calibration, SEC),但其只能评价模型的有效性。因此,验证集的SEP是评价模型的重要统计学参数。其计算方法如下:

$$SEP=\sqrt{\frac{\sum_{i=1}^{n}(y_{v,i}-Y_{v,i})^2}{n}}$$

其中,$y_{v,i}$,指化学参考值;$Y_{v,i}$;指验证集近红外预测值;n为样本数。

$$SEC=\sqrt{\frac{\sum_{i=1}^{n}(y_{c,i}-Y_{c,i})^2}{n-p}}$$

其中,$y_{c,i}$,指校正集化学参考值;$Y_{c,i}$;指校正集近红外预测值;n为样本数;p为变量数(对于MLR而言,是选用的波长数;对于PCR而言,是主成分数;对于PLS而言,是潜变量因子数)。

在建立与优化模型时,如果不采用内部验证,那么独立的外部验证集是必不可少的。

(4)化学参考值的测定。

近红外定量分析方法的性能取决于参考方法的精密度和准确度。因此,应尽可能使参考方法的不确定度降至最低。应对参考方法的重复测定结果进行讨论,并以图表形式给出参考数据。应依据参考方法和近红外方法的精密度与准确度,对参考分析方法的重复测量次数进行说明,并证明其合理性。

3. 模型校正

(1)软件。

采集完校正集样本的近红外光谱和化学参考值之后,需将每个样本的光谱值与其参考值对应,然后进行化学计量学建模。可采用特定的分析软件进行建模。这类软件主要是凸显数据间的差异性,并将差异进行关联。对于不同的算法,其数据关联结果的表现形式不同(例如,对于PLSR为潜变量因子数,而对于PCA为主成分数)。

(2)主成分的选择。

为避免数据欠拟合或过拟合,校正模型中主成分数(或其它能够表征数据差异性的参数)的选取至关重要。

在选取校正模型的主成分数时,应充分考虑近红外分析方法的适用范围,以及校正集样本对市售产品的代表性。同时应考虑以下几点:共线性;待测组分光谱变化对数据变异的最低贡献;由样本基质中非待测组分(如辅料或其它成分)所导致的数据差异。

实际工作中,还应考虑由可行性研究显露出来的以及与待测组分性质相关的所有因素。用以描述由每个主成分所解释的变异的载荷图可有助于主成分数的选择。对于不涉及主成分的算法(如神经网络),对其关键参数的选择应给出解释说明。

一旦选定主成分数,可用图表或统计学参数对相应校正模型进行表征,其特征参数为校正标准差(SEC)或其它类似的参数。偏差和截距等参数也可用于模型评价。

(3)模型的优化。

建立近红外分析方法时,最佳校正模型的选择至关重要。校正模型可采用内部验证进行优化(见"校正集,测试集及验证集样本组成"),这些方法也可用于评价校正模型正确预测定量结果的适用性。

目前,采用重抽样统计方法进行模型优化是一

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

个快速发展的领域，且已发展出更多适用的统计方法，尤其是用以评价数据欠拟合或过拟合的方法。对所采用的任何方法都应进行解释说明并证明其合理性。

采用内部（交叉）验证进行模型优化时，其特征参数通常是交叉验证标准差（SECV）。其它统计学参数如偏差或变异系数等也可用于模型评价。对模型优化过程中所涉及的任何统计学参数都应进行说明并证明其合理性。

$$SECV=\sqrt{\frac{\sum_{i=1}^{n}(y_{cv,i}-Y_{cv,i})^2}{n}}$$

其中，$y_{cv,i}$，指化学参考值；$Y_{cv,i}$，指近红外预测值；n 为样本数。

4. 模型验证

（1）概述。

为了证明近红外方法的线性、范围和专属性，需采用独立的（外部）验证集样品。

在对未知样本进行测定时，近红外定量分析方法依赖于参考分析方法获取的参考值或已知的样本组成，因此，需采用某一统计学标准对模型的预测能力进行评价。对（外部）验证集样本而言，评价参数为预测标准差（SEP）。

近红外定量分析方法应采用下述参数进行验证。

（2）预测标准差。

对于外部验证集，应计算其 SEP 和 SEP/range（参考化学值范围），同时结合其它统计参数，如表征参考方法精密度的实验室标准差（Standard Error of Laboratory, SEL）。这些参数应依近红外光谱分析方法的目的进行相应的解释、讨论与评价。

$$SEL=\sqrt{\frac{\sum_{i=1}^{n}(y_{1,i}-Y_{2,i})^2}{n}}$$

其中，$y_{1,i}$ 与 $y_{2,i}$，指不同实验条件下测定的化学参值；n 为样本数。

（3）专属性。

近红外光谱分析方法的专属性取决于它应用的目的、科学对象与范围。

所谓专属性是指所建方法能甄别其适用范围以外的样本，如不合格样品、空白样品、辅料配比不同的样品和含有不同活性成分和辅料的样品等（参见一般要求中“光谱质量统计检验”）。

以下各项可为专属性提供支撑：证明待测组分光谱测量合理性的可行性研究的数据；建模所用主成分的载荷和待测组分 NIR 特征吸收之间的比较；证明方法准确性和稳健性的验证数据。

（4）线性。

为测定方法的线性，验证集样本的浓度应分布在模型适用浓度范围内。否则，不能充分地确认和验证方法的线性。如条件允许，应将近红外光谱法的预测结果与参考分析方法进行对比，并将预测值和参考值的相关系数和残差分析（线性表征）的结果以图表的形式展现，并进行解释说明。

在测定线性时，若采用《指南》以外的参数，应做出适当说明。

（5）范围。

范围应通过涵盖模型适用浓度范围的验证集进行确认。验证集样本的浓度若超出校正集浓度范围，其在近红外光谱分析时应表现为离群值。

（6）准确度。

对于近红外方法，在其特定适用范围内应测定其准确度，且准确度应适于应用目的。

（7）精密度。

精密度和 SEP 应根据近红外光谱分析方法的应用目的来设定合理的值。偏差应尽可能接近 0。近红外光谱分析方法的精密度是否恰当，应结合待测组分予以充分讨论和解释。

光谱重复性取决于样本的装样条件，可通过重复测量光谱测定其重复性。整个样本变异范围的光谱重复性均应进行考察。

中间精密度可通过不同时间不同分析者的测定结果进行求算。

（8）稳健性。

一般而言，近红外定量分析中参考分析方法只测定样本的化学特性或物理特性，而近红外光谱法可同时测定样本的物理和化学特性。

依据所建方法的目的和样本的取样条件，验证方法稳健性时应同时考虑化学和物理影响因素，包括温度、湿度、样本取样差异和仪器变化等（见一般要求中“影响光谱响应的因素”）。可考虑采用 DOE 使有效信息最大化。

在整个近红外分析过程中应解决方法稳健性问题。若提前考虑到这点，那么模型构建和优化过程中产生的数据和验证数据便足以证明方法的稳健性。否则，需要另外提供测定和验证方法稳健性

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的验证数据。

(9) 检出限与定量限。

必要条件下或待测成分是杂质时，需要测定近红外光谱分析方法的检出限和定量限。

七、近红外光谱分析方法的生命周期与批准后要求

1. 近红外光谱分析方法生命周期的维护

近红外光谱分析方法（定性或定量分析）会随时间改变（见图 1），因此，需在最初应用时说明如何对近红外光谱分析方法进行维护。

应明确界定近红外光谱分析方法的适用范围（见一般要求中“近红外光谱分析范围的确定”），这对后续如何对其进行改进至关重要。只有在经过变更申请后才能对近红外的适用范围进行扩展。

2. 已认证近红外光谱分析方法的变更

对于影响近红外光谱分析方法性能的任何变更（有计划的和无计划的），都需对近红外模型进行重新校正和/或验证，以证明模型的适用性。任何变更都应按照有效的变更管理协议进行相应的验证，并记录在案。

(1) 近红外光谱分析方法范围内的变更。

一般情况下，近红外光谱分析方法界定范围内的变更只需进行 GMP 检查。如光谱库的维护、仪器配件（包括光源、采样装置、仪器位置）的更换以及软件升级等。任何变更都应记录在案，并涵括相应的再验证报告和兼容性报告，以表明修订后的近红外光谱分析方法与批准程序相一致。还需对变更进行风险评估。

对于近红外光谱定性分析方法，每一方法及其参考光谱库需进行适宜的变更管理测试。这一测试最少应包含两个样本集（即两个等级或两种物质的样本集），变更后的方法须完全区分这两个样本集。如果变更后的方法不符合变更管理测试（即变更方法无法区分这两个样本集），则需对模型进行重新验证。变更管理测试的适用性也需定期进行评估。

近红外定量分析方法只能按照模型校正时的条件，在校正浓度范围使用。在进行模型更新时，可适当向校正模型中添加新样本（但必须在近红外光谱分析方法的适用范围内）。模型变更后需要进行重新验证，并对变更记录在案以便 GMP 检查。

(2) 近红外光谱分析方法范围外的变更。

为了更新参考光谱库，须为近红外方法适用范围的扩增提供变更申请，申请中应包括更新后方法与当前方法和/或参考方法的比较。

在扩大近红外光谱定性分析方法的应用范围时，如在光谱库中添加新的样本，应说明其与《指南》的兼容性，并将更新后的方法与当前方法进行对比。

在扩大近红外光谱定量分析的应用范围时（如改变校正浓度范围和/或规范要求），需要对近红外模型重新进行校正和验证。

3. 仪器间近红外光谱分析方法的传递

近红外光谱分析方法传递的目的是确保在一台仪器上建立的校正模型能够在其它仪器上使用（判断依据是验证参数，详见定性和定量分析方法中“模型验证”）。同一样品在“主仪器”（建立模型所用仪器）和其它“被传递仪器”（接受模型传递的仪器）上所采集光谱的预测结果应相同。

判断仪器相似性的基本参数包括：硬件（如光谱仪类型和附件）；软件（包括算法和校正模型中光谱的处理方法）；光源-样品接触面（如光纤探头和光纤）。

根据近红外分析方法的应用范围（如定性分析、趋势分析或定量分析）及所用仪器间的相似程度，可采用下列措施之一进行模型传递。

(1) 仅当近红外光谱采集方法改变时。

可应用近红外光谱仪自带软件中的数学补偿，以确保代表性样本集在不同仪器上获得的光谱响应相同。一般可采用偏差校正、斜率校正或矢量校正。

(2) 数学补偿不理想的所有其他情况。

应在被传递仪器上重新进行模型校正及验证。建立传递模型时，可选取一部分代表性的样本分别在主仪器和被传递仪器上进行光谱采集。可依据多元杠杆作用（Multivariate leverage）进行代表性样本的选择，多元杠杆作用越大，样本对校正模型的影响越大，越具有代表性。依据模型的复杂程度不同，所选取的少量代表性样本（少于初始用于校正和验证模型的样本数）应足以支持模型在仪器间的传递。在主仪器不可用的情况下，应采用合理数量的代表性样本在被传递仪器上建立校正模型。

上述两种情况下，近红外分析方法在不同仪器间

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的传递均需遵从可比性方案(Comparability protocol)。

如果对后续变更有所预见,可将模型传递的可比性方案作为“批准后变更管理协议”纳入到初始申请书中。还可在变更申请时递交可比性方案的相关资料,以注册传递模型。可比性方案中应包含经验证合理的模型传递成功与否的判定标准。

Introduction to the “Guideline on the Use of Near-infrared Spectroscopy by Pharmaceutical Industry and Data Requirements for New Submissions and Variations (Draft)”

Qiao Yanjiang, Du Min, Shi Xinyuan, Wu Zhisheng, Xu Bing

(School of Chinese Pharmacy, Beijing University of Chinese Medicine, Beijing 100102, China)

Abstract: This paper was aimed to introduce a draft guideline newly issued by the European Agency for the Evaluation of Medicinal Products (EMEA) which is intended to provide guidance to pharmaceutical industry in the process of calibration, validation and maintenance of a near-infrared spectroscopy (NIRS) method when used for quantitative/qualitative analysis and process monitoring and the type of data requirements for new submissions and variations.

Keywords: Near-infrared spectroscopy, process analytical technology

(责任编辑:李沙沙 张志华,责任译审:王 晶)

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net