

JJF

中华人民共和国国家计量技术规范

JJF XXXX—202X

检验医学标准物质互换性评估规范

Evaluation Criteria of Commutability of Reference Materials

for Laboratory Medicine

(征求意见稿)

202×-××-××发布 202×-××-××实施

国家市场监督管理总局发布

检验医学标准物质互换性 评估规范

JJFXXXX-202X

Evaluation Criteria of Commutability of Reference Materials for Laboratory Medicine

本规范经国家市场监督管理总局于 202×年××月××日批准，并
自 202×年××月××日起施行。

归口单位：全国标准物质计量技术委员会

主要起草单位：中国计量科学研究院

参加起草单位：北京航天总医院

国家卫生健康委临床检验中心

广东省中医院

上海市临床检验中心

本规范委托全国标准物质计量技术委员会负责解释

本规范主要起草人：

参加起草人：

国家标准物质计量技术委员会规范征求意见稿

目录

引言.....	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
3.1 互换性	1
3.2 基体效应	2
4 互换性评估原理	2
5 互换性验证临床样本的选择	2
5.1 单供体来源临床样本	3
5.2 多供体来源临床样本	3
5.3 临床样品的添加	3
6 互换性验证评价方法	3
6.1 互换性验证的线性回归数学模型	3
6.2 戴明回归分析	4
7 互换性的报告	6
附录 A 实例	8
附录 B：基于 IFCC 方案的偏倚差值法	11

引言

检验医学标准物质互换性是标准物质作为测量量值溯源与传递的载体的重要特性，标准物质互换性的评价规范对于确保上述活动的有效性具有重要意义。本规范的主要目的是规范检验医学标准物质的研制过程，为医学检验等活动提供符合质量要求的标准物质。

本规范在参考GB/T 19703-2020《体外诊断医疗器械生物源性样品中量的测量有证参考物质及支持文件内容的要求》、GB/T 21919-2008《检验医学参考测量实验室的要求》、WS/T 356《基质效应与互换性评估指南》、CLSI EP30-A《医学检验用标准物质互换性的特性值和条件》、CLSI EP14-A3《处理过样品的互换性评估》等国内外规范性技术文件的基础上。Commutability 对应的中文包括互换性及互通性，本规范中统一使用“互换性”；Reference Material (RM) 对应的中文包括标准物质、参考物质或标准样品，本规范中统一使用“标准物质”。

国际临床化学和检验医学联合会（IFCC）互换性评价工作组于2018年发布了最新的互换性评价方案建议，该方案通过计算不同测量程序中标准物质的偏差与临床样品的偏差之间的一致性来评估互换性。本规范以附录形式给出，仅作为参考提供。

本规范为首次发布。

检验医学标准物质互换性评估规范

1 范围

本规范规定了检验医学标准物质互换性的评价方法，标准物质包括以下物质：有证标准物质、参考品、校准品等具有标准物质特性的物质，以及室间质评（EQA）或者能力验证（PT）的样品，标准物质的均匀性、稳定性不在本规范的考察范围内，本规范所表述的内容默认为待考察标准物质已符合均匀性、稳定性的要求。

注：本规范亦可为管理部门、认可机构和其他机构确认或识别标准物质研制（生产）机构的能力提供参考依据。

2 规范性引用文件

JJF 1001 通用计量术语及定义

JJF 1005 标准物质通用术语和定义

GB/T 19703-2020 体外诊断医疗器械生物源性样品中量的测量有证参考物质及支持文件内容的要求（ISO 15194: 2009 IDT）

GB/T 21919-2008 检验医学参考测量实验室的要求（ISO 15195: 2003 IDT）

WS/T 356 基质效应与互换性评估指南

CLSI EP14-A3 处理过样品的互换性评估（Evaluation Of Commutability Of Processed Samples）

CLSI EP30-A 用于检验医学可互换标准物质的特征和限定条件；已批准的指导手册（Characterization and Qualification of Commutable Reference Materials for Laboratory Medicine）

凡是注日期的引用文件，仅注日期的版本适用于本规范；凡是不注日期的引用文件，其最新版本（包括所有的修订本）适用于本规范。

3 术语和定义

JJF1001、JJF1005、GB/T19703-2020 中规定的被测量、标准物质、不确定度、测量模型、测量程序、有证标准物质等术语及以下术语适用于本规范。

3.1 互换性 commutability

标准物质的特性。将标准物质和能够代表预期测量样品类型的样品均采用不

同测量程序测量，该特性由所得测量结果之间数学关系的等效性证明。【ISO Guide 30 2.1.20】

注 1：该定义采自 CLSI EP30-A 可互换标准物质的定值与鉴定（Characterization and Qualification of Commutable Reference Materials）。

注 2：参见 JJF1001-2011、ISO/IEC 指南 99：2007 和 GBT21415-2008。

注 3：在一些领域里，互换性又称作互通性。

3.2 基体效应 matrix effect

除被测量以外，样品特性对特定测量程序测定被测量及其量值的影响。【GB/T 21415 3.15】

注 1：某个基体效应的明确原因即为一个影响量。

注 2：“基体效应”有时被错误用于因被分析物的变性或加入非真实组分（代用品）以模拟分析物等缺少互换性。

注 3：“基体效应”有时又称“基质效应”。

4 互换性评估原理

互换性评估是用不同测量方法同时对选定的一系列具有代表性的临床样本和候选标准物质进行分析，利用临床样本的测量结果建立数学关系，观察候选标准物质的测量结果偏离这一数学关系的程度，偏离越大，说明候选标准物质在所用方法之间的互换性越差。

两种测量方法之间的数学关系很大程度上依赖于所选择的待测样品的性质。由于临床检验分析的对象是临床样本，所以在本规范中应使用具有代表性的临床样本作为比对标准。

当用标准物质进行检验医学溯源或校准时，应针对该测量程序对标准物质进行互换性评估。检验医学标准物质的互换性仅存在于测量方法与比对方法之间，当任意一个方法发生变化时，互换性需重新评估。进行互换性评估前应设计互换性评估方案，方案中应明确被测量、测量方法、比对方法、实验样品、统计方法等内容。应该在相同的条件下测量临床样本和待评价的标准物质，通过数学相关性确定互换性。

5 互换性验证临床样本的选择

用于互换性评估的临床样本应该能够覆盖标准物质的浓度范围，并且不少于 25 份。

临床样本除非收集后可以立即用于测试，一般情况下需要分装并保存在-70℃以下的环境中，应用前避免反复冻融。

5.1 单供体来源临床样本

互换性验证过程中，优先使用单供体来源的临床样本。通常从健康人或者病人的个体收集临床样本。

5.2 多供体来源临床样本

在收集数量不足的情况下，可以选择多供体混合临床样本。在混合样本的过程中，由于个体的非特异性相互作用，可能会造成沉淀，干扰互换性结果。

因此在混合样本前需要考察被测量的稳定性，尽量缩短耗时，避免反复冻融样本，若无法避免，应验证冻融过程对分析物稳定性的影响。

5.3 临床样品的添加

不建议为了获得高浓度样本而向临床样本或者样本池中添加提纯或部分提纯的分析物，并且避免使用添加剂(例如抗生素或者其它防腐剂)。添加剂或者添加剂中的杂质可能会对基体产生不良影响，使样品的性质发生改变。必须添加分析物时，要估算临床样本的近似值。当使用添加剂时，需说明这些替代临床样本对互换性结论产生的潜在局限性。

6 互换性验证评价方法

互换性验证需要在不同的测量程序间，建立相应的数学模型。验证互换性的主要方法是线性回归。在建立数学模型前，应确定数据是否符合正态分布，需采用合适的方法检验离群值并剔除。

IFCC 互换性评价工作组推荐的偏倚差值法 (difference in bias) 使用临床预期用途相关的固定标准来量化标准物质与临床样品的接近程度，具有临床应用价值。该方案通过计算不同测量程序中标准物质的偏差与临床样品的偏差之间的一致性来评估互换性。该方法评价原理及方法列于附录 B，仅作为参考提供。

本章 6.2 为互换性评价的线性回归方法提供具体指导。计算实例参见附录 A。

6.1 互换性验证的线性回归数学模型

线性回归方法假设在两个变量 (x , y) 之间存在一种线性关系，估算线性参数和标准差。通常使用戴明回归方法、普通最小二乘法 (OLS)、和 Passing-Bablok

(PB) 方法。戴明回归方法根据观察曲线与回归线之间的最小均方误差和, 估算斜率和偏移量。OLS 通常假设变量 y 常数方差的分布误差, 自变量 x 为无误差自变量。PB 回归方法适用于两个变量都不精确的情况, 是一个非参数化过程, 对误差项的分布类型没有任何假设, 可以用于常数项或比例误差项。当数据满足线性关系时, 戴明回归分析方法是评价互换性的最佳方法。

在测量误差理想化的情况下, 使用两种测量方法对临床样品进行测量, 得出的两种测量结果 y_H 和 x_H , 可以绘成一条直线。如果代表标准物质测量结果的 y_R 和 x_R 点, 位于上述直线上, 认为标准物质与临床样本具有相同的数学关系, 则标准物质与应用同一对测量方法进行测量的临床样本可互换; 反之, 不可互换。

6.2 戴明回归分析

采用 X、Y 两种测量程序分别对临床样本和标准物质进行测定, 假设测量结果之间存在线性关系, 临床样本和标准物质测量结果之间的关系表示为:

$$y = \alpha_H + \beta_H x \quad (1a)$$

$$y = \alpha_R + \beta_R x \quad (1b)$$

α_H, α_R : 临床样本和标准物质分别测定, 得到一对结果值之间的线性关系的截距;

β_H, β_R : 临床样本和标准物质分别测定, 得到一对结果值之间的线性关系的斜率。

临床样本和标准物质测量结果之间的数学关系 (1a) 和 (1b) 在各自的模型参数成对相等时等价:

$$\alpha_H = \alpha_R \quad (2a)$$

$$\beta_H = \beta_R \quad (2b)$$

两种测量方法中都存在测量误差时, 临床样本的模型 (1a) 表示为:

$$y = \alpha_H + \beta_H(x + \varepsilon_X) + \varepsilon_Y \quad (3)$$

α_H : 临床样本测定得到一对结果值之间线性关系的截距;

β_H : 临床样本测定得到一对结果值之间线性关系的斜率;

ε_X : X 测量程序中的随机误差;

ε_Y : Y 测量程序中的随机误差。

多次重复测量时, 模型 (3) 可表示为:

$$\bar{y}_i = \alpha_H + \beta_H(\bar{x}_i + \varepsilon_X) + \varepsilon_Y \quad (4)$$

α_H : 临床样本测定得到一对结果值之间线性关系的截距;

β_H : 临床样本测定得到一对结果值之间线性关系的斜率;

ε_X : X 方法的测量平均值的随机误差;

ε_Y : Y 方法的测量平均值的随机误差。

假设标准物质与临床样本在 X 方法和 Y 方法之间可互换, 则标准物质 Y 方法的测量结果 y_R 与 X 方法的测量结果 x_R 呈线性关系:

$$y_R = \alpha_H + \beta_H x_R \quad (5)$$

α_H : 临床样本测定得到一对结果值之间线性关系的截距;

β_H : 临床样本测定得到一对结果值之间线性关系的斜率;

y_R 、 x_R 表示标准物质分别用 Y 方法、X 方法测量得出的结果。

在模型 (5) 中经过 r_R 重复的 Y 方法测量的平均结果的预测值 $\bar{y}_{R_pred}$ 受模型参数估计误差和测量随机误差的影响表示如下:

$$\bar{y}_{R_pred} = \hat{\alpha}_H + \hat{\beta}_H(\bar{x}_R + \varepsilon_{\bar{x}_R}) + \varepsilon_{\bar{y}_R} \quad (6)$$

$\hat{\alpha}_H$: 临床样本测定得到一对结果值之间线性关系的截距;

$\hat{\beta}_H$: 临床样本测定得到一对结果值之间线性关系的斜率;

$\bar{x}_R$: 标准物质的 x 变量的平均值;

$\varepsilon_{\bar{x}_R}$: X 方法测定标准物质时测量平均值的随机误差;

$\varepsilon_{\bar{y}_R}$: Y 方法测定标准物质时测量平均值的随机误差。

$\bar{y}_R$ 的 $(1 - \gamma)100\%$ 预测区间计算方式为:

$$\bar{y}_{R_pred} \pm t(1 - \gamma/2, \nu) \hat{\sigma}(\bar{y}_{R_pred}) \quad (7)$$

ν : 预测值 $\bar{y}_{R_pred}$ 的标准差 $\hat{\sigma}(\bar{y}_{R_pred})$ 的自由度;

γ : 显著性水平; 对于传统使用的具有 95% 置信度的统计区间, $\gamma = 0.05$ 。

通过在图表上绘制 $\bar{y}_R$ 和 $\bar{x}_R$ 点, 建立标准物质的可互换性, 画一条直线说明以下方程式:

$$y = \hat{\alpha}_H + \hat{\beta}_H x \quad (8)$$

$\hat{\alpha}_H$: 临床样本测定得到一对结果值之间线性关系的截距;

$\hat{\beta}_H$: 临床样本测定得到一对结果值之间线性关系的斜率。

按上述方法计算并绘制 95% 的预测区间。当 $\bar{y}_R$ 和 $\bar{x}_R$ 点落在预测区间内时, 标

准物质视为可互换；反之，视为不可互换。当标准物质具有系列浓度时，应对每个浓度都进行互换性评估。

使用线性回归方法进行互换性评价的结果图，如图 1 所示。如果假设通过方法 X 和方法 Y 测量的每个临床样本和候选标准物质都进行相同的测量次数并得到平均值，并且在每种方法的测量范围内方差恒定的，那么这两个程序测定的临床样本测量结果之间的回归模型的 95% 预测区间表示基于 r 次测量结果的平均值将落在 0.95 的概率范围。类似地，可互换的候选标准物质基于 r 次测量结果的平均值的应与临床样本表现相同，其测量值应以 0.95 的概率落在这些范围内。如果上述候选标准物质结果的平均值落在临床样本的 95% 预测区间之外，则该候选标准物质被认为是不可互换的。

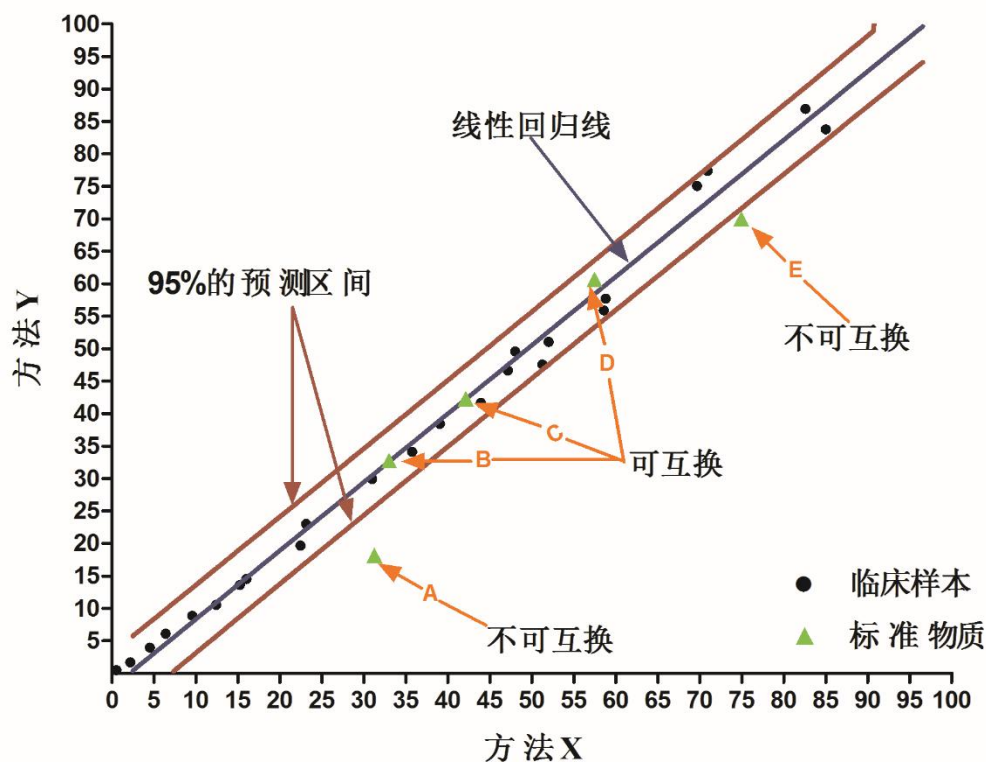


图 1 方法 X 和方法 Y 的互换性评估结果

图 1 显示在方法 X 和方法 Y 中，三个候选标准物质（B、C、D）可以与临床样本互换，因为根据临床样品的结果显示，它们的结果都包含在 95% 的预测区间内，两个候选标准物质（A、E）是不可互换的，因为它的结果在预测区间之外。

7 互换性的报告

互换性报告应包括以下信息：

- 1) 用于互换性验证的临床样本描述;
- 2) 在互换性评估中使用的参考测量程序或测量程序 (包括测量方法及试剂信息)
- 3) 用于建立临床样本标准物质数学关系的描述。
- 4) 标准物质的互换性结果。标准物质对于某个测量程序的不可互换, 反映了该标准物质不适用于该特定测量程序的校准或验证。

互换性评定结果表示中需要列出在互换性研究中使用的所有成对测量程序的结果。通过为每个标准物质创建交叉表格进行记录。如表 1 所示。

在单元格中, 如果物质具有互换性, 输入 1; 如果物质不具有互换性, 输入 0。当无法判断时, 输入-1。交叉表格有助于评估物质的互换性和帮助终端用户 (例如: 体外诊断试剂厂家或者 PT/EQA 提供者) 确定该标物是否适用于指定用途。

表 1. 互换性结果表示 (交叉表格)

	方法 A	方法 B	方法 C	方法 D	方法 E
方法 A		1	0	-1	0
方法 B			1	1	1
方法 C				0	1
方法 D					1
方法 E					

附录 A 实例

C-反应蛋白

ERM-DA474/IFCC、GBW09867 两种需要评估互换性的标准物质，相对扩展测量不确定度分别为 6.1%、4.1%；

25 例临床血清；

结果如下：

表 A.1 ID-LC-MS/MS 和免疫比浊法对 C-反应蛋白的测量结果

样本编号	ID-LC-MS/MS 法均值 (mg/L)	免疫比浊法均值 (mg/L)
血清 1	0.53	0.37
血清 2	2.18	2.17
血清 3	4.53	4.73
血清 4	6.40	6.77
血清 5	9.55	9.73
血清 6	12.40	12.20
血清 7	15.23	15.50
血清 8	15.99	16.73
血清 9	22.44	22.27
血清 10	23.13	25.60
血清 11	30.99	31.87
血清 12	35.73	36.37
血清 13	43.94	43.67
血清 14	39.04	39.80
血清 15	47.14	47.47
血清 16	51.26	47.83
血清 17	48.01	49.50
血清 18	51.98	50.33
血清 19	58.78	56.13
血清 20	58.55	55.17
血清 21	69.66	68.23
血清 22	70.91	70.67
血清 23	82.57	80.53
血清 24	84.99	78.00
血清 25	97.53	96.90
ERM-DA474/IFCC	41.79	41.13
GBW09867	51.60	53.37

以 ID-LC-MS/MS 方法测量结果为 X 轴、免疫比浊法测量结果为 Y 轴，绘制两法的线性回归曲线和参考方法测量结果的 95%置信区间。

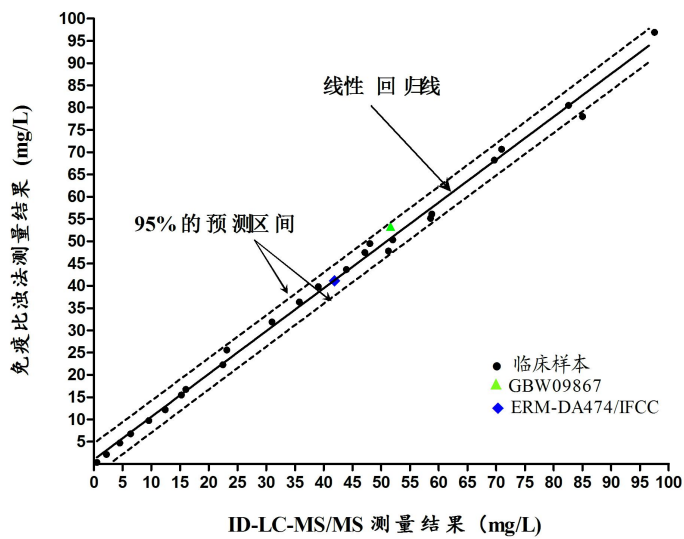


图 A.1 使用 95%的预测区间对 C-反应蛋白进行互换性评估结果

总胆红素

A、B、C、D 四种需要评估互换性的标准物质，相对扩展测量不确定度为 3%

26 例临床血清

结果如下：

表 A.2 参考方法和常规方法对总胆红素的测量结果

样本编号	参考方法均值 ($\mu\text{mol/L}$)	常规方法均值 ($\mu\text{mol/L}$)
血清 1	4.96	4.10
血清 2	11.07	12.00
血清 3	16.84	18.33
血清 4	24.66	27.27
血清 5	37.00	41.03
血清 6	39.13	51.57
血清 7	42.49	48.87
血清 8	54.11	59.83
血清 9	59.67	53.77
血清 10	65.33	60.73
血清 11	66.65	61.30
血清 12	68.19	61.90
血清 13	77.14	71.07
血清 14	92.80	91.13
血清 15	105.21	81.03
血清 16	118.27	112.43
血清 17	132.32	138.93
血清 18	136.58	117.63

血清 19	141.09	122.23
血清 20	142.72	129.43
血清 21	215.34	211.00
血清 22	335.32	316.70
血清 23	356.93	305.90
血清 24	553.67	535.67
血清 25	638.37	603.40
血清 26	679.73	725.33
A	108.10	108.50
B	44.60	43.57
C	53.16	56.30
D	200.70	218.23

以参考方法测量结果为 X 轴、常规方法测量结果为 Y 轴，绘制两法的线性回归曲线和参考方法测量结果的 95%置信区间。

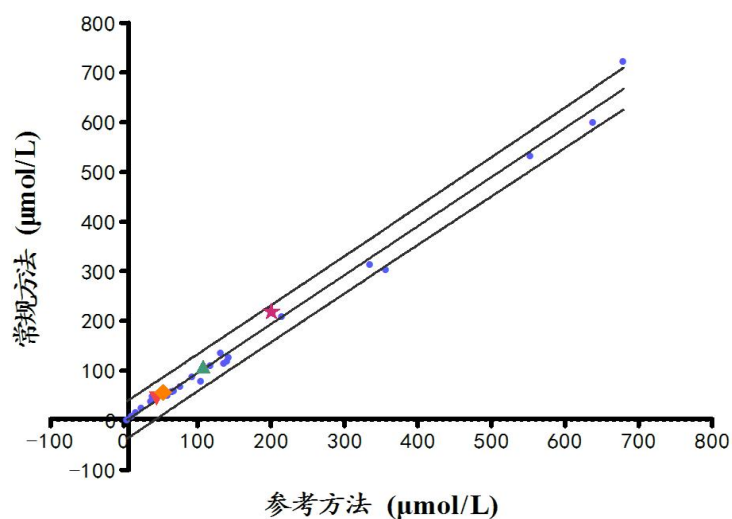


图 A.2 使用 95%的预测区间对总胆红素进行互换性评估结果

附录 B：基于 IFCC 方案的偏倚差值法

IFCC 推荐的偏倚差值法（difference in bias）使用临床预期用途相关的固定标准来量化标准物质与临床样品的接近程度，具有临床应用价值。通过偏倚差值法中的误差模型，可以估计方法之间偏倚的变异来源，进而更好的解释互换性的评价结果。本附录就基于 IFCC 方案偏倚差值法的互换性评价方案作一介绍。

注：该方法参考 IFCC 发表的以下文献：[1] ClinChem, 2018,64(3):447-454. [2] ClinChem, 2018,64(3):455-464. [3] ClinChem, 2018,64(3):465-474.

1. 偏倚差值法理论思想

1.1 误差模型

偏倚差值法互换性评价基于误差模型（error model）理论，误差模型模拟了不同测量程序测量结果差异的误差组分。在典型的 IFCC 互换性评价方案中，通常使用一定量的临床样本在一个批次内用不同测量程序进行若干次重复测定，其中两个测量程序测定单个临床样本的测定结果差异的简易模型如公式 B.1 所示。

$$y - x = b(\mu) + d + e_y - e_x, \quad (\text{B.1})$$

其中：

y ：该临床样本方法 y 测定值；

x ：该临床样本方法 x 测定值；

μ ：该临床样本的真实浓度；

$b(\mu)$ ：方法 y 与方法 x 测定结果之间的常见偏差，可为恒定常数或 μ 的连续函数；

d ：样本特异性效应，可认为是在人群中选择样本的随机变异组分；

e_x ：方法 x 的批内变异组分；

e_y ：方法 y 的批内变异组分。

$b(\mu)$ 可表示为 μ 的连续函数，意味着样本浓度 μ 发生微小变化会引起 $b(\mu)$ 发生变化。术语 d 指样本特异性效应，而组分 d 的标准差示为 σ_d 。若其中一个或两个测量程序的选择性提高，可减小样本特异性。测量程序批内变异 e_x 和 e_y 的标准差分别示为 σ_x 和 σ_y 。理想情况下，测量程序的批内变异是完全随机的， σ_x 和 σ_y 可通过重复测量的结果进行评估， σ_d 通过随机变异估计值排除批内变异组分而得。若不满足方法间偏差完全来源于随机变异的前提条件，那么公式 B.1 的误差组分模型还需包含位置效应组分。而临床样本的重复测量如果不是连续测量，在评估 σ_d 时会将位置效应变异也计算在内。其中位置效应是指测量样本时由于不同检测位置带来的变异，是批内方法性能的一个趋势性变异组分。在互换性评价中，建议所有样本检测在一个测量批内完成，并且对同一个临床样本

表 B.1 实际情况下方方法间偏差可能存在的变异来源

变异类型	可能来源	评估方式
重复测量变异	一批内连续重复测量带来的变异	通过样本连续重复测量的结果评估
位置效应变异	一批内不同位置带来的变异，批内方法性能的一个趋势性的变异组分	通过设置标准物质不同位置组来评估
样本特异性变异	来源于在人群抽样的变异，不同样本可能有不同的性质（比如基体效应）	通过比较随机变异与批内变异来确定
浓度趋势性变异	来源于不同浓度引起的方法间偏差的趋势性的变化	通过比较观察到的变异与随机变异来确定

进行连续重复测量，其中位置效应通过在不同位置组测量标准物质来进行评估。 $b(\mu)$ 和 σ_d 通常与选取的临床样本盘有关，但其标准差假定与浓度趋势近似无关。所有误差组分的分布假定为近似正态分布。表 B.1 汇总了方法间偏差可能存在的变异来源和评估方式，并在下文的误差组分评估中进行了详细介绍。

1.2 互换性评价

偏倚差值法互换性评价是通过在标准物质的浓度下，评估标准物质（RM）在两个方法之间的偏差与临床样本（CS）在两个方法之间的偏差是否一致。所有临床样本方法间偏差的均值记作 B_{CS} ，标准物质的方法间偏差记作 B_{RM} ，偏倚差值 d_{RM} 为 B_{CS} 和 B_{RM} 的差值（式 2）。为了评定标准物质的互换性，需根据指定偏倚差值 $|d_{RM}|$ 的允许范围，即互换性评价限（C），互换性评价限通常根据标准物质的临床用途确定，如最大允许偏倚。当测量值是 k 个重复测量的平均值时，随机变异 e 和 d 对方法间偏差变异的贡献 σ_{Random} 如式 3。

$$d_{RM} = B_{RM} - B_{CS}, \quad (B.2)$$

$$\sigma_{Random} = \sqrt{\sigma_d^2 + \frac{\sigma_x^2}{k} + \frac{\sigma_y^2}{k}}, \quad (B.3)$$

而标准物质的互换性通过比较偏倚差值 d_{RM} 、偏倚差值扩展不确定度 $U(d_{RM})$ ，与互换性评价限 C 的关系来评定，分为以下三种情况：

- （1）若 $d_{RM} \pm U(d_{RM})$ 在 $0 \pm C$ 之内，则表示标准物质具有互换性；
- （2）若 $d_{RM} \pm U(d_{RM})$ 在 $0 \pm C$ 之外，则标准物质不具有互换性；
- （3）若 $d_{RM} \pm U(d_{RM})$ 在 $0 \pm C$ 范围重叠，则表示标准物质的互换性无法判定。

当 $U(d_{RM}) > C$ 时，无法评估标准物质的互换性，过大的 $U(d_{RM})$ 可能是由于不合适的实验设计（重复测量次数或临床样本数量不够），也可能是检测方法

的低精密度和低选择性造成样本特异性效应过高引起的。若是实验设计的不足，应更改实验设计并重新实验。若是检测方法的原因，应剔除性能不佳的测量程序，因此在进行互换性实验之前，需提前了解各个测量程序的精密度及选择性，通过资格预审以纳入互换性评价。

2. 互换性评价实验设计

偏倚差值法互换性实验的设计方案应该实现评估偏倚差值及偏倚差值的扩展不确定度，同时评估方法间偏差的变异组分以确定变异的主要来源。因此，推荐的实验设计方案如下。收集 n 个临床样本，参与互换性评价的测量方法分别在一批之内测定所有临床样本，每个临床样本连续重复测定 k 次。临床样本的测定顺序需随机排序，避免按照浓度排序。每个待评价的标准物质分成 p 组，代表 p 个不同的位置，穿插分布在整批测定之间，并在相应位置上连续重复测定 k 次。若已知其他因素造成方法间偏差，则在实验设计中需考虑该因素的影响。例如，测量程序使用多通路取样针添加样本或试剂，则每个通路需进行相同次数的重复测量，即 k 至少等于通路数目。

临床样本数目 n 至少需要 30 个，且其浓度需在整个区间内均匀分布，才能区分样本特异性效应以及与浓度相关的系统偏差 $b(\mu)$ 。临床样本浓度范围无需覆盖每个测量程序的测量区间，但需覆盖每个标准物质及相邻浓度范围。重复测量次数 k 最小为 2 次，但出于精密度离群值剔除的考虑，推荐进行 3 次重复测量。若测量程序的精密度为主要变异来源，则需要的重复测量次数更多。通过测量不同位置的标准物质可以评估位置效应，并对标准物质与临床样本的偏倚差值扩展不确定度进行更好的评估。位置数目 p 推荐不低于 5，则可以满足位置效应评估时所需的自由度。互换性度量的估计值， d_{RM} ，必须在可接受的不确定度范围内确定。作为最低要求，建议 $U(d_{RM}) < C/2$ 。若使用 n 、 k 和 p 的推荐最小值无法满足不确定度的要求，该实验设计能够识别不确定度的主要来源，并指出后续实验需要的改进方向。

3. 数据初步分析

互换性评价数据处理时使用差分图进行统计分析，可以区分浓度趋势对变异组分造成的影响。在统计分析中，随机误差组分需与浓度无关。若不能满足该要求，可对数据进行自然对数转换（ln 转换）或其他转换，ln 转换的优势在于可以

更好地解释结果。是否进行数据转换和数据转换形式需根据数据初步分析结果来决定。

进行互换性评价之前，每个测量程序需进行精密度考察，如图 B.1 所示，计算该测量程序每个临床样本重复测量的标准差和平均值并作图。如图 B.1A 所示，如果该方法的精密度考察未表明其标准差与浓度之间存在强相关（斜率大于 2），则可合并所有临床样本的标准差作为该方法重复测量标准差的估计值，记作 s_y 。若临床样本的标准差与浓度成比例关系（图 B.1B），则结合差分图考虑是否进行数据转换。

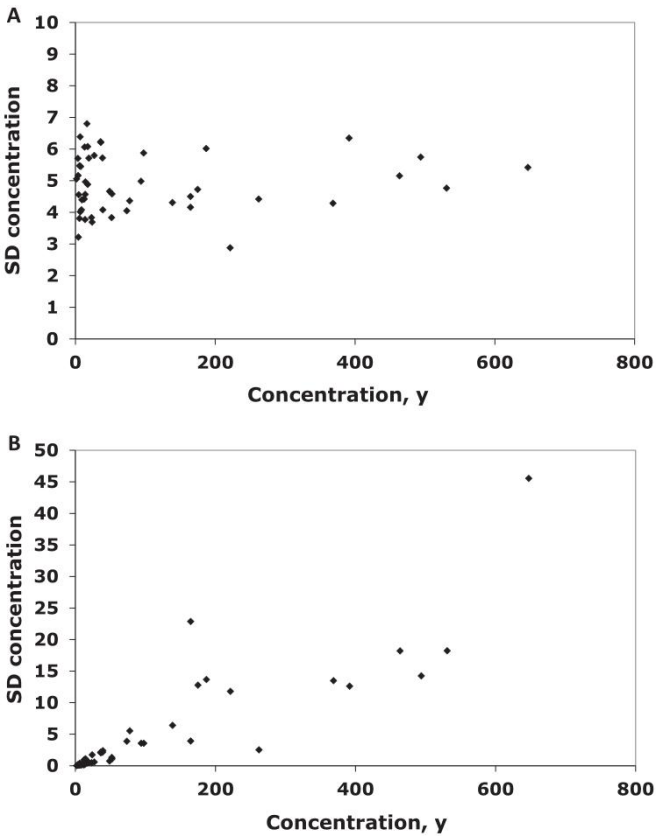


图 B.1 精密度考察

注：A 表明区间内临床样本重复测量的标准差为常数，B 表明标准差与浓度呈比例关系。

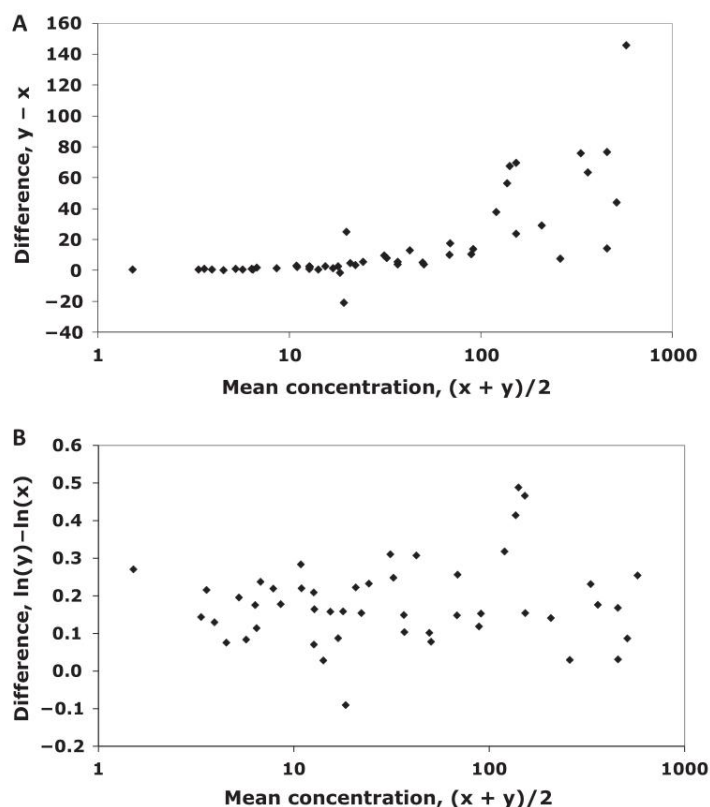


图 B.2 未进行数据转换 (A) 和自然对数转换后 (B) 的差分图

差分图的纵坐标为两个测量程序检测临床样本的差值或数据转换后的差值, 当有参考测量程序时横坐标为参考测量程序的测定值, 反之使用两个测量程序测定值的均值。差分图的分布宽度受到所有随机组分变异的影响, 通过对差分图进行视觉检查, 观察其分布与浓度相关程度来决定是否使用数据转换。图 B.2 为使用原始数据和数据转换后的差分图, 明显可以观察到当其分布与浓度相关时, 更适合使用数据转换。与精密度考察相比, 差分图的分布与浓度不存在比例关系更为重要, 是数据处理中首要满足的前提条件。若数据转换后仍不能满足差分图分布在该浓度区间近似恒定, 则应在差分图分布近似恒定的不同浓度区间分别对标准物质进行互换性评价。

对精密度和差分图分布进行考察后, 需对离群值进行剔除。当离群值相对较少时 ($<10\%$), 可直接对离群值进行剔除避免对评价结果造成影响。视觉检查通常足以实现离群值的识别, 从精密度考察和差分图中可识别出明显的离群值。如果某个数据难以判断, 可以在不剔除和剔除该数据的情况下分别对数据进行分析。若不剔除该潜在离群值对评价结果没有明显影响, 则可以包含该数据。

4. 误差组分评估

4.1 源于标准物质的变异组分

源于检测标准物质的变异组分主要为位置效应变异和重复测量变异。在互换

性评价时,对于每个标准物质设置了 p 个位置组,并且在每个位置重复连续测量了 k 次。可以通过单因素方差分析来判断每个测量程序是否存在位置效应,其检验统计量 F 如式 B.4 所示。位置组均值间的标准差记作 $s_{Pos-mean}$,为该测量程序检测标准物质总变异的估计。而 $s_e/\sqrt{k}$ 为该测量程序重复测量变异的估计,其中 s_e^2 为不同位置组组内方差的均值。若方差分析结果判定存在位置效应,则位置效应变异 s_{Pos} 的评估如式 B.5 所示。当 $F < 1$,即式 B.5 的根号内为负数时,可将 s_{Pos} 估计值设为数值零。以上为单个标准物质的误差组分计算,若互换性评价存在多个标准物质时,可对标准物质的变异组分分别进行合并,即所有标准物质变异均方差的平方根。注意在合并 s_{Pos} 时,不能使用数值零替代根号内的负数进行计算。

$$F = \frac{s_{Pos-mean}^2}{s_e^2/k}, \quad (B.4)$$

$$s_{Pos} = \sqrt{s_{Pos-mean}^2 - \frac{s_e^2}{k}}, \quad (B.5)$$

4.2 源于临床样本的变异组分

4.2.1 评估重复测量变异

重复测量变异通过合并所有临床样本重复测定的标准差而评估,根据不同的测量程序分别记作 s_y 和 s_x 。若标准物质的重复测量变异与临床样本的重复测量变异通过方差检验未发现明显差异,则标准物质的重复测量变异可以包括在 s_y 和 s_x 中,否则应分别对标准物质和临床样本使用各自的重复测量变异估计值。

4.2.2 评估浓度趋势性变异

临床样本方法间偏差即差分图的分布可由两个部分组成,第一个部分是拟合到散点中心的连续函数,以及第二部分函数分布的变异。第一部分即为 $b(\mu)$ 的估计值,第二部分应该为随机变异组分即公式 3 中的 σ_{Random} 。理论上来说,连续浓度临床样本的方法间偏差 B_i (式 B.6) 应具有近似相同的预期均值,可以从每对连续浓度下方法间偏差中估计偏差函数周围的变异。合并所有连续浓度下方法间偏差 B_i 的差值得出基于均方递差 (MSSD) 的变异估计值 s_{MSSD} (式 B.7)。式 B.7 中的 B_i 按照浓度的升序排序,该变异估计值 s_{MSSD} 包括了样本特异性组分,排除了浓度趋势性变异的组分,通常作为 σ_{Random} 的估计值。其中观察到的临床样本间方法偏差变异即为 B_i 间的标准差,记作 s_B (式 B.8)。如方法间偏差函数为常数即无浓度

趋势性变异存在, s_{MSSD} 与 s_B 应不存在较大差异。因此, 对 s_{MSSD} 和 s_B 进行方差检验, 若 s_{MSSD} 远小于 s_B , 则证明有浓度趋势性变异。

$$B_i = y_i - x_i, \quad (B.6)$$

$$s_{MSSD} = \sqrt{\frac{1}{2(n-1)} \cdot \sum_{i=1}^{n-1} (B_{i+1} - B_i)^2}, \quad (B.7)$$

$$s_B = \sqrt{\frac{1}{(n-1)} \cdot \sum_{i=1}^n (B_i - B_{CS})^2}, \quad (B.8)$$

y_i : 临床样本 i 方法 y 测定均值;

x_i : 临床样本 i 方法 x 测定均值;

n : 临床样本数量;

B_i : 临床样本 i 的方法间偏差。

B_{CS} : 所有临床样本方法间偏差的均值。

4.2.3 评估样本特异性变异

如上所述, s_{MSSD} 是对随机变异组分 σ_{Random} 的估计值。因此当不存在样本特异性效应时, s_{MSSD} 的估计值应如式 B.9 所示, 即重复测定变异。对 s_{MSSD} 和 $\sigma_{Random(0)}$ 进行方差分析, 其检验统计量 F 如式 B.10, 若 s_{MSSD} 远大于重复测定变异, 则说明存在样本特异性变异。

$$\sigma_{Random(0)} = \sqrt{\frac{\sigma_x^2}{k} + \frac{\sigma_y^2}{k}}, \quad (B.9)$$

$$F = \frac{k \cdot s_{MSSD}^2}{s_x^2 + s_y^2}, \quad (B.10)$$

则样本特异性变异估计值 s_d 的计算方式如式 B.11, 若无浓度趋势性变异存在, 应该用 s_B 替换式 B.11 中的 s_{MSSD} 。若在估计标准物质的变异组分中发现测量程序存在位置效应, 则应对式 B.11 的样本特异性变异 s_d 进行校正, 校正后的样本特异性变异估计值 $s_{d(corr)}$ 如式 B.12。

$$s_d = \sqrt{s_{MSSD}^2 - \frac{s_x^2 + s_y^2}{k}}, \quad (B.11)$$

$$s_{d(corr)} = \sqrt{s_{MSSD}^2 - \frac{s_x^2 + s_y^2}{k} - s_{Pos(x)}^2 - s_{Pos(y)}^2}, \quad (B.12)$$

无论是否存在样本特异性效应, 互换性评估和扩展不确定度 $U(d_{RM})$ 的计算以相同的方式进行。然而, 若不确定度太大从而无法对标准物质互换性进行判定, 则必须确定在不确定度中起主导作用的误差分量, 从而改进互换性评估实验方案并重新评估。具体改进措施例如, 是否应该增加重复检测的次数或临床样本的数

量，是否应该剔除某些临床样本，或者因某个测量程序的选择性不佳，是否在互换性评价中排除将该测量程序等。

5. 互换性评价

标准物质的互换性评估通过标准物质在两个方法之间的偏差与临床样本在两个方法之间的偏差的差值 d_{RM} 及其扩展不确定度 $U(d_{RM})$ 来进行评定的。其中不确定度 $u(d_{RM})$ 由源于标准物质的不确定度和源于临床样本的不确定度两部分构成，其中来源于标准物质的不确定度评估如式 B.13 所示。其中 $s_{Pos-mean(x)}$ 和 $s_{Pos-mean(y)}$ 分别为在源于标准物质变异组分评估中方法 x 和方法 y 的位置组均值间标准差。

$$u(B_{RM}) = \sqrt{\frac{s_{Pos-mean(x)}^2 + s_{Pos-mean(y)}^2}{p}}, \quad (B.13)$$

B_{RM} 的不确定度 $u(B_{RM})$ 是确定的，而 $b(\mu)$ 在标准物质浓度下的估计值和不确定度取决于互换性的实验结果，可能存在以下四种情况：

- (1) 偏差函数 $b(\mu)$ 在整个临床样本浓度区间恒定（即常数），且覆盖了标准物质的浓度。

在此情况下临床样本偏差的估计值为所有临床样本方法间偏差的均值 B_{CS} ，其不确定度 $u(B_{CS})$ 的评估如式 B.14。 d_{RM} 为 B_{RM} 和 B_{CS} 的差值（式 B.2），其不确定度 $u(d_{RM})$ 的评估如式 B.15。

$$u(B_{CS}) = \frac{s_B}{\sqrt{n}}, \quad (B.14)$$

$$u(d_{RM}) = \sqrt{\frac{s_{Pos-mean(x)}^2 + s_{Pos-mean(y)}^2}{p} + \frac{s_B^2}{n}}, \quad (B.15)$$

- (2) 偏差函数 $b(\mu)$ 在包含标准物质的浓度区间近似为常数。

若在整个浓度区间存在浓度趋势变异，则找到一个覆盖标准物质浓度的区间（ q 个临床样本），在此区间内偏差函数 $b(\mu)$ 不受浓度因素影响。在此情况下， q 的数目必须足够大，以给出可接受的不确定度，其不确定度的 $u(d_{RM})$ 与情况（1）类似，不同之处在于用 q 替换 n 进行计算（式 B.16）。

$$u(d_{RM}) = \sqrt{\frac{s_{Pos-mean(x)}^2 + s_{Pos-mean(y)}^2}{p} + \frac{s_B^2}{q}}, \quad (B.16)$$

- (3) 在部分浓度区间内（ q 个临床样本），偏差函数 $b(\mu)$ 是一个关于浓度的连续

函数，且标准物质的浓度在该浓度区间的中间位置。

样本数目 q 必须是足够大的偶数，以给出可接受的不确定度，临床样本也需在该浓度区间内均匀分布。用 q 个临床样本方法间偏差的均值作 B_{CS} 的估计值，其源于临床样本的不确定度的评估需要使用 s_{MSSD} 替代 s_B 进行计算，来排除浓度趋势性的影响（式 B.17）。若偏差函数存在一个明显的浓度性趋势，可能无法进行互换性评价。

$$u(d_{RM}) = \sqrt{\frac{s_{Pos-mean(x)}^2 + s_{Pos-mean(y)}^2}{p} + \frac{s_{MSSD}^2}{q}}, \quad (B.17)$$

(4) 不满足上述 (1)、(2) 和 (3) 三种情况。例如，只有少量甚至不存在浓度接近于标准物质的临床样本。

在相关浓度区间内没有临床样本或临床样本数目很少的情况下，无法验证该模型的合理性，也无法得出可靠地互换性评价结果。

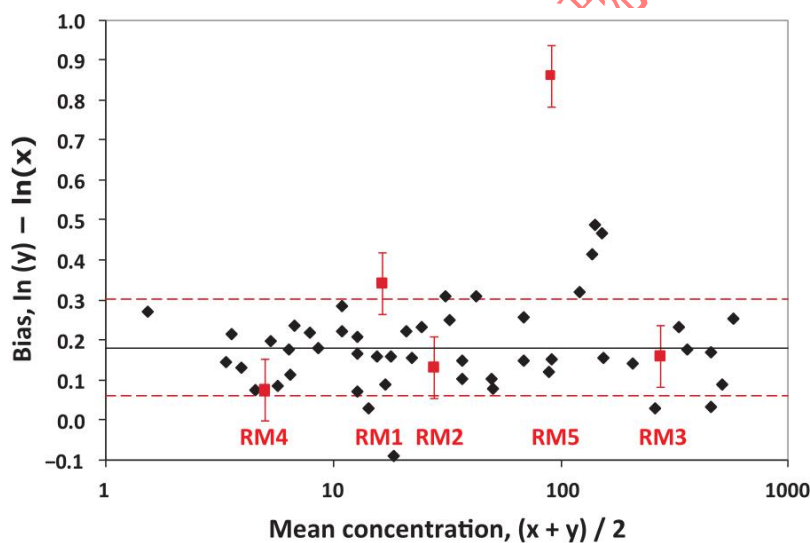


图 B.3 偏倚差值法进行互换性评价示意图

最终，标准物质的互换性评价结果展示如图 B.3。黑色实线为临床样本的方法间偏差均值 B_{CS} ，红色虚线为互换性评价限，黑色方块为每个临床样本的方法间偏差，红色方块为标准物质的方法间偏差 B_{RM} ，方块上的红色实线为偏差差值的扩展不确定度 $U(d_{RM})$ 。如图所示，RM1、RM2 和 RM4 的互换性评价结果为无法判定，RM3 的结果为具有互换性，而 RM5 的评估结果为不具有互换性。