

包装试验方法确认指南

Standard Guide for Packaging Test Method Validation

草案稿

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX - XX - XX 发布

XXXX - XX - XX 实施

目 次

前言	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 意义和用途	3
4.1 实验室间研究	3
4.2 属性试验方法确认(ATMV)	5
4.3 变量试验方法确认(VTMV)	11
附录 A (资料性) 可接受准则示例	18
附录 B (资料性) ATMV 工作表示例	19
参考文献	20
图 1 观察变异	12
图 2 观察变异	12
表 1 ATMV 示例	5
表 2 选择 VTMV 方法和可接受准则操作注释	13
表 3 Gage R&R 研究建议的最小试验次数	14
表 A.1 可接受准则示例	18
表 B.1 ATMV 工作表示例	19

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国医疗器械行业协会提出。

本文件由中国医疗器械行业协会医疗器械包装专业委员会归口。

本文件起草单位：通标标准技术服务（上海）有限公司、上海微创医疗器械（集团）有限公司、亚都控股集团有限公司、南微医学科技股份有限公司、东莞市安保医用包装科技有限公司、杜邦（中国）研发管理有限公司。

本文件主要起草人：童玉玺、李然、许慧、丁艳琴、王燕、秦蕾、李婧秋、夏飞、赵欣。

本文件首次发布于2022年12月。

引 言

在诸如医疗器械或制药等受监管行业中，工程师会经常使用一些运用广泛的试验方法，这些方法在国家标准和行业标准中都有引用。然而，围绕这些试验的确认问题并没有被很好地理解。经常出现的问题是：这些试验方法是否应该得到确认？如何确认它们？应该确认到什么程度？

对此GB/T 19633.1和GB/T 19633.2提供的答案是：“所有用于表明符合本部分的试验方法应得到确认，并形成文件。”不过，这个申明在如何证明符合这些需求方面几乎没有提供答案。这是由于在如何使用这些试验方法方面需要有很大的灵活性，并不是所有的情况和试验方法都需要同样程度的审查。因此，在评估什么时候、为什么以及如何确认试验方法时，牢记这种灵活性，并针对既定的情况使用可用的最佳工具，来适当地回答上述问题是至关重要的。一个健全的风险评估过程，可以说是确定与被试验的特定设计要素相关的风险的最佳工具。例如，从患者安全的角度来看，标签附着力试验与无菌屏障完整性试验的风险存在明显差异。如果标签丢失，产品将被丢弃，并选择一个贴有正确标签的新产品。然而，如果无法检测到无菌屏障包装密封破裂或材料上的针孔，则患者可能会因为使用受污染的器械而感染。

确定与医疗器械包装组件相关的风险级别的典型过程是失效模式与影响分析工具，通常称为FMEA。FMEA过程旨在确定产品或过程的潜在失效模式，评估与这些失效模式相关的风险，根据重要性对问题进行排序，并确定和记录解决最严重问题的缓解策略，有许多指南和标准都描述了这一过程¹⁾。本指南将有助于提出一些方法来解决：基于相关风险，什么试验方法确认（TMV）方式在给定应用中是最有效的，并为确认的执行提供指南。

1) SAE J1739, AIAG FMEA-3和MIL-STD-1629A。

包装试验方法确认指南

1 范围

本文件提供的信息旨在阐明使用包装试验方法的机构，以及通过实验室间研究（ILS）²⁾确认包装试验方法的过程，解决机构的实验室间研究的共识标准以及特定的方法。

本文件为定义和开发计量型和计数型数据应用的方法确认提供了指南。

本文件适用于医疗器械及包装行业开展包装试验方法确认活动。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 3358.2-2009 统计学词汇及符号 第2部分：应用统计

GB/T 19633.1-2015 最终灭菌医疗器械包装 第1部分：材料、无菌屏障系统和包装系统的要求

YY/T 1759-2020 医疗器械软性初包装设计与评价指南

T/CAMDI 058-2020 最终灭菌医疗器械包装——GB/T 19633.1和GB/T 19633.2应用指南

ASTM F17 初级屏障包装术语

ASTM E177 ASTM试验方法中精度和偏差术语的使用规范

ASTM E456 有关质量和统计的术语

ASTM E691 进行实验室间研究以确定试验方法的精确度的实施规程

ASTM E2282 定义试验方法试验结果的指南

ASTM E2782 测量系统分析指南（MSA）

3 术语和定义

GB/T 3358.2-2009界定的以及下列术语和定义适用于本文件。

3.1

准确度 accuracy

是指测试结果或测量结果与真值之间的一致程度。

3.2

α 风险错误（ α ） alpha risk error (α)

检验员拒绝合格单元的概率，也被称为生产方风险或第一类错误。在本文件中，这种错误类型被称为 α 风险错误或（ α ）。

3.3

评定员 appraiser

用于识别执行试验方法确认活动的人员，通常也被称为鉴定员或技术员。

3.4

属性试验方法 attribute test method

对符合或不符合的质量特性合格或失败的输出试验。某些情况下变量测量数据也可作为属性测量数据进行处理。

3.5

β 风险错误（ β ） beta risk error (β)

检验员接受不合格单元的概率，也被称为 β 错误（漏判率）或第二类错误。在本文件中，这种错误类型被称为 β 风险错误或（ β ）。

2) 实验室间研究（ILS）讨论的重点是编写和解释试验方法精密度声明，以及分析和说明结果的替代方法。

3.6

临界样品 borderline samples

处于合格或不合格临界边缘的样品。

3.7

比较检验方法 comparative test method

一种使用统计检验对两个或两个以上总体的均值进行比较的检验方法（如双样本t检验、方差分析检验）。比较检验方法不用于接受或拒绝单个部件，输出通常没有规范限值。

3.8

失效模式与影响分析 failure modes and effects analysis**FMEA**

是一种逐步识别设计、制造或装配过程或产品或服务中所有可能失效的方法³⁾。

3.9

批允许缺陷水平 lot tolerance percent defective**LTPD**

在抽样计划中，表示抽样计划中对孤立批规定的常规拒收的质量水平。等于或高于LTPD的批次将在由置信度确定的概率水平被拒绝。LTPD可由抽样计划的操作特性曲线（OC曲线）确定。也称为不合格质量水平（RQL）、极限质量水平（LQ）和不可接受质量水平（UQL）。

3.10

测量分辨率 measurement resolution

又称分辨力，指测量系统识别并显示被测量物最微小变化的能力。

3.11

精密度 precision

在规定条件下，多次独立测试/测量定结果之间互相接近的一致程度。

3.12

精密度与公差比（%P/T）（测量能力指数） precision to tolerance ratio（%P/T）

%P/T是一个Gage R&R研究的试验方法性能指标。它测量可归因于试验方法变异公差的百分比。根据所评估的检测方法变异的组成部分，%P/T有三种形式：%P/T重复性、%P/T再现性和%P/T总和。

3.13

重复性 repeatability

同一操作员按照相同的方法，使用相同的测试或测量设施，在较短时间间隔内对同一测试/测量对象进行观测获得测试/测量结果的精密度。

3.14

再现性 reproducibility

不同的操作员按照相同的方法，使用不同的测试或测量设施，对同一测试/测量对象进行观测获得独立测试/测量结果的精密度。

3.15

再现性和重复性误差百分比（%R&R）

%R&R是Gage R&R研究的试验方法性能指标。它测量归因于试验方法变异的历史过程变异百分比。计算%R&R需要已知的历史标准偏差。

3.16

自证 self-evident

满足以下两项标准的检验：

——不合格的本质是离散于标准，这意味着合格与不合格之间不存在被误会的过渡区间。

——只需要做少量的培训或者不做培训就可以区分合格与不合格。

3.17

%研究变异（%SV） % study variation（%SV）

Gage R&R研究的试验方法性能指标。它测量由试验方法变异引起的Gage R&R研究总变异的百分比。

3) <http://asq.org/learn-about-quality/process-analysis-tools/overview/fmea.html>。

3.18

行业专家 subject matter expert (SME)

产品和/或工艺的行业专家。对失效模式有深刻理解的工程师、检验员、技术员、培训师和生产主管可被视为行业专家。

3.19

测试系统 test systems

执行试验所需的仪器和相关材料。

3.20

试验 trial

试验定义为一名检验员或一台设备，在使用高度仪器化方法进行一次测量或做出通过/失败的判断。如果三名检验员对同一设备进行一次评估，则计为三次试验。同样，如果一名检验员在试验期间对同一设备进行了两次评估，则将其计为两次试验。

3.21

确认团队 validation team

试验方法确认的责任方，寻求跨职能输入，确认试验方法的有效性，并完成与任何试验失败相关的纠正措施。

3.22

变量试验方法 variable test method

一种关于连续标度产生数值结果的试验方法。

4 意义和用途

4.1 实验室间研究

4.1.0 通则

通过实验室间研究和特定的方法来寻求机构的共识标准。为了能够使试验结果可靠，在很多情况下，试验方法需要被确认。开展试验的机构应完成这些确认，在通过实验室间研究制定标准时也要实施确认，但不能代替检测机构进行的确认工作。

4.1.1 检测机构的确认

检测机构的确认包括计划、执行和分析研究。计划应包括对试验方法适用范围的描述，包括试验设备、样本的测量范围、样本的选择原理、样本量以及方法学选择的依据。

4.1.2 实验室间研究的目的

实验室间研究重点不是试验方法的开发，而是在成功完成开发阶段后收集用于试验方法精密度说明所需的信息。实验室间研究获得的数据可能表明，需要进一步改进试验方法。在这种情况下，精密度被定义为试验方法的重复性和再现性，通常称为量具 R&R（也称Gage R&R）。对于实验室间研究，每个实验室都要有自己独立的试验系统。如果实验室间研究以这种方式进行，同一实验室的评定员和试验系统之间的再现性是不被评估的。

4.1.3 实验室间研究过程概述

实验室间研究过程包括计划、执行和分析，旨在评估试验方法的精密度的研究。做到这一点所需的步骤是：创建任务组，确定实验室间研究协调员，创建试验设计，执行试验，分析结果，并在试验方法中记录结果精密度的声明。有关如何进行实验室间研究的更多细节，见ASTM E691。

4.1.4 编写精密度和偏倚声明

本文件的精密度和偏倚声明的编写，可参见ASTM E177中概述的标准操作规程。在某些情况下，以表格形式呈现信息，标准使用者更易理解，示例见4.1.5。

4.1.5 分析和陈述结果的替代方法——计量型数据

4.1.5.1 能力研究

- a) 大于 2.00 的过程能力表明试验输出的总变异（零件间变异和试验方法变异）相对于公差非常小。数学上，

$$Pp = \frac{\text{Specification Tolerance}}{6 \sigma_{\text{Total}}} \geq 2.00 \quad \dots\dots\dots (1)$$

$$\sigma_{\text{Total}} \leq \frac{1}{12} \text{Specification Tolerance}$$

式中：

Specification Tolerance——规范公差；

Pp ——过程性能指数，也称长期能力指数。强调从过程总波动的角度考察过程能力；

σ_{Total} ——过程总波动， $\sigma_{\text{总变异}}$ ；

σ_{Part} —— $\sigma_{\text{零件间变异}}$ ；

σ_{TM} —— $\sigma_{\text{方法变异}}$ 。

- b) 上式中 σ_{Total} 包含 σ_{Part} 和 σ_{TM} ，可以得出两个结论：
 1) 试验方法可以区分至少 1/12 的公差，因此试验方法的分辨率是足够的，不必要做额外的分析，如 Gage R&R 研究。
 2) 测量相对于规范公差，测量是精确的。
 c) 由于试验方法确认的能力研究需要两个或多个操作员使用一个或多个测试系统，高能力指数将证明试验方法在不同操作员和不同测试系统试验间的一致性。

4.1.5.2 Gage R&R 研究

- a) 以下针对 %SV、%R&R 和 %P/T 的建议验收标准来自全行业对测量系统采用的要求。如果试验方法变异（ σ_{TM} ）小于研究总变异的 30%（ σ_{Total} ）则可以接受该试验方法⁴⁾。
 b) 等效于：一个大于 2.00 的过程能力表明试验输出的总变异（零件间变异和试验方法变异）相对于公差非常小。数学上，

$$\%SV = \frac{\sigma_{\text{TM}}}{\sigma_{\text{Total}}} \leq 30\% \quad \dots\dots\dots (2)$$

- c) 当有历史数据来评价过程的变异时，还应满足：

$$\%R\&R = \frac{\sigma_{\text{TM}}}{\sigma_{\text{Process}}} \leq 30\% \quad \dots\dots\dots (3)$$

式中：

σ_{Process} —— $\sigma_{\text{过程变异}}$

- d) 对于 %P/T，另一个惯例是使用正态分布中间 99% 的分布范围来表示“全”分布⁵⁾。理想情况下，输出的公差范围应该大于这个比例。对于正态分布的总体，这表明：

$$\text{Specification Tolerance} \geq 5.15 \sigma_{\text{Total}} \quad \dots\dots\dots (4)$$

- e) 上式中因子 5.15 为正态分布的双边 99% 置信的西格玛水平 Z 值。因此：

$$\%P/T = \frac{\sigma_{\text{TM}}}{\text{Specification Tolerance}} \leq \frac{\sigma_{\text{TM}}}{5.15 \times \sigma_{\text{Total}}} \leq \frac{30\%}{5.15} = 5.8\% \quad \dots\dots\dots (5)$$

式中：

Specification Tolerance——规范公差

- f) 实际上，这意味着一个达到 6% P/T 再现性的试验方法在评估给定设计的 P/T 时是有效的。

4.1.5.3 功效和样本量研究

当使用统计检验比较两个或多个总体的平均值时，过多的试验方法变异性可能掩盖真实的差异，并降低统计检验的能力。因此，需要一个大的样本量来维持统计检验足够的检验功效（ $\geq 80\%$ ）。当从业

4) 汽车工业行动集团(AIAG)测量系统分析手册(第4版, 第78页)。

5) Burdick, Borror和Montgomery的 Gage R&R研究的设计和分折, 第3页。

务的角度来看，样本量太大而无法接受时，应在运行比对试验之前改进试验方法。因此，可根据比对试验方法对其功效和样本量的影响（如双样本t检验）来决定是否接受该方法。

4.2 属性试验方法确认（ATMV）

4.2.1 属性试验方法确认的目标

属性试验方法确认（ATMV）证明，向检验员提供的培训和工具使他们能够高度成功地区分符合的产品和不符合的产品。有两个标准用来衡量ATMV是否达到了这一目标：首要准则是证明最大漏判率 β 小于或等于其规定的 β_{\max} 阈值；第二准则是证明最大误判率 α 小于或等于其规定的阈值 α_{\max} 。

4.2.2 ATMV 流程概述

在属性试验方法确认中，实施的是一个由合格单元和不合格单元组成的单盲研究。当两种抽样方案均满足时，ATMV通过。第一个抽样方案表明试验方法满足最大允许 β 错误（漏判率）的要求，第二个抽样方案表明试验方法满足最大允许 α 错误（误判率）的要求，即试验方法能够证明它以高水平的有效性接受合格单元和拒绝不合格单元。 β 错误抽样计划完全由不合格单元组成，每个检验员⁶⁾进行的 β 试验的总数被汇总在一起，它们的错误分类（被接受的不合格单元）总数需要小于或等于 β 错误抽样计划规定的失败数量。 α 错误抽样计划完全由合格单元组成，每个检验员所进行的 α 试验的总数被汇总在一起，并且他们错误分类（被拒绝的合格单元）的总数需要小于或等于 α 错误抽样计划所规定的失败数量。

4.2.3 ATMV 示例

属性试验方法涵盖了广泛的试验范围。表1列出了这些试验方法类别的示例。

表1 ATMV 示例

检验员	变量测试数据的试验方法组成		属性测量数据的试验方法组成	
	触觉	视觉	触觉	视觉
人类	针规	尺寸模板	已完工表面的突起或毛刺	产品的气泡、缝隙或变色
机器	用于试验印刷电路板的钉床	自动化成像系统	接触式轮廓仪	自动检验系统

4.2.4 计量型数据的 ATMV

一般通过量化测量数据分析计量型试验方法。某些情况下，测量数据转化为计数型数据时会更有效。例如：用于医疗器械的无菌屏障系统（SBS），其密封强度规范要求为（4.45-6.67）N/25.4mm，这个数值可被测量。当使用分辨力只有0.04N的拉力试验机来测量密封强度时，即使在生产中很少出现不符合的产品，其最小过程性能指数（Ppk）还是很差。此时，ATMV比计量型方法确认更适合，方法确认团队宜将数据作为属性进行确认。

4.2.5 自证检验

自证检验的要求要满足缺陷在本质上是离散的，几乎不需要或根本不需要培训就能检测。

4.2.5.1 以下情况可视为自证检验

- 当导线上的润滑水平正确时，传感器指示灯就会亮起，否则当润滑不足时就熄灭。试验设备正在为检验员创建一个二进制输出，并且说明很简单，这是自证的。请注意，试验方法确认涉及的设备也需要进行确认。
- 组装中的组件——如果组件的存在相当容易被检测到，这是自证的，因为结果是二进制的。
- 组装中使用的正确组件——只要组件彼此不同，这是自证的，因为结果是二进制的。

4.2.5.2 以下通常不被视为自证检验

6) 检验员可能是一台机器。

- a) 燃烧或热变色--除非组件在过热时完全变色，否则该检验将要求检验员检测变色的痕迹，这不能满足离散条件的要求。
- b) s 弯或 z 弯成型不当--将部件放置在模板顶部，检验员确认构件完全在模板边界内。弯曲可以在多个位置从完美形状到完全超出边界，在每一级弯曲之间的变化。因此，这不是一个离散的结果。
- c) 组件表面无划痕——划痕的大小从“放大后不可见”到“肉眼不可见”到“肉眼可见”不等。因此，这不是一个离散的结果。
- d) 组件表面无毛刺——由于检验员的手指上有老茧，他们的触摸灵敏度不同，毛刺的锋利度和暴露度也不同。因此，这既不是一个离散的条件，也不是一个易于培训的指令。
- e) 组件有裂纹——裂纹的长度和严重程度各不相同，检验员看到视觉缺陷的能力也各不相同。因此，这既不是一个离散的结果，也不是一个易于训练的指令。

4.2.6 ATMV 步骤

4.2.6.1 步骤 1

准备试验方法文件：

- a) 确保在 ATMV 执行之前，设备鉴定已经完成或至少在确认计划中完成。
- b) 试验方法文档中要记录的设备设置示例，包括环境或环境条件、显微镜放大倍数，自动检测系统照明和进给速度，真空衰减试验压力和洁净室的照明标准，这可能涉及在洁净室里获取照度计读数来表征光线水平。
- c) 与培训人员合作创建缺陷图片，包括符合产品的图片和不太极端的缺陷示例可能是有益的，因为示例的范围将为决策提供更好的分辨力。
- d) 在可能的情况下，宜采用与检验过程中一致的放大倍数来显示视觉设计标准。
- e) 确保 ATMV 使用最新的视觉设计标准运行，并能很好地反映潜在缺陷。

4.2.6.2 步骤 2

建立可接受准则：

- a) 识别哪些缺陷需要包含在试验中。
- b) 使用报废历史记录来识别每个缺陷代码或类型的频率。这也可以是由行业专家简单提供的信息。
- c) 一个检验步骤中不宜设置太多缺陷。随着更多的缺陷被添加到检验过程中，检验员最终将到达无法检验所有内容的地步，这个阈值也可能在 ATMV 试验中显示出来。限值因产品类型和试验方法而异，但对于目视检验，15-20 是可能达到的最大缺陷数量。

4.2.6.3 步骤 3

确定每个缺陷所需的性能水平：

- a) 如果 ATMV 试验先于风险分析完成，建议的方法是使用最坏情况结果或高风险确定。这需要与由于更保守的评级带来的样本量的增加进行权衡。
- b) 没有相关风险指数的失效模式可以根据确认团队同意的任何要求进行试验。如果某个部件或组装件可能因为某种特定的失效模式而报废，良好的商业意识是通过执行 ATMV 来确保检验是有效的。
- c) 由于低分辨率和严格的规范限值，针规是一个变量测量数据有时被当作属性数据处理的示例。在这种应用中，检验员在试验前要接受培训，以理解可接受和不可接受分别对应的分歧水平。
- d) 来料检验是经常将变量测量数据视为属性数据的另一个示例。将变量测量作为通过/失败的结果来处理，可以允许使用不太复杂的测量工具（如模板），并且对检验员要求的培训也更少。然而，宜将受益与可能需要的更多样本以及信息丢失的程度进行权衡比较。例如，属性数据无法区分位于规范限中值的样本和恰好位于规范限两边的样本，这可能会导致更高的下游成本并使提高产量过程中解决问题的难度更大。

4.2.6.4 步骤 4

确定可接受准则：

- a) 参考公司预定义的置信度和可靠性要求；
- b) 参考附录 A 中的图表示例。

4.2.6.5 步骤 5

创建确认计划，确定样品中各缺陷的比例：

- a) 虽然宜就缺陷比例如何在 ATMV 中分配提供某种理论依据，但在选择比例方面有一定的灵活性。因此，不同的产品和工艺可能采用不同的策略，例如 30 个零件中有 10 个缺陷零件或 30 个零件中有 20 个缺陷零件。样本的成本以及与错误结果相关的风险会影响决策。
- b) 新产品往往没有废品生产数据。在这种情况下，使用来自类似产品的历史数据，或者根据开发期间观察到的过程挑战估计预期的废品比例。另一种选择是平均地分配所有缺陷。

4.2.6.6 步骤 6

确定所需检验员和设备数量：

- a) 试验次数很多时，可考虑聘请 3 名以上检验员，以减少试验所需的独特部件的数量。更多的检验员可以检验相同的部件而不需要增加更多的部件，以实现额外的试验和更高的统计功效。
- b) 不要求检验员全部检验相同的样品，虽然这可能是最简单的方法。
- c) 对于需检验员放置夹具或设置敏感的半自动检验系统，仍宜雇用多名检验员进行试验。
- d) 对于完全独立于检验员的自动检验系统，只需要一名检验员。然而，为了减少所需独特部件的数量，可以考虑控制其他变量，如各种照明条件、温度、湿度、检验时间、白班/夜班和部件方向。

4.2.6.7 步骤 7

在试验之前对检验员进行培训：

- a) 向检验员说明 ATMV 的目的和重要性。
- b) 检验员培训应为双向过程。确认小组应就检验文档中视觉标准、图片和书面描述的质量和清晰度征求检验员的反馈。
 - 1) 是否有任何需要澄清的灰色区域？
 - 2) 图表是否会比缺陷的实际图片更有效？
- c) 审查临界样品。考虑将临界样品的图片/图表添加到视觉标准中。在某些情况下，功能缺陷和外观缺陷之间可能会有所不同。这可能因方法/包装类型而异。
- d) 有些确认小组已经进行了空运行试验，以表征当前检验的有效性。如果空运行试验和最终试验都涉及相同的检验员，则不宜使用相同的样品。

4.2.6.8 步骤 8

选择具有代表性的检验员组作为试验组：

- a) 某些情况下，如现场转移，所有检验员对产品的熟悉程度大致相同。如果是这种情况，则根据检验员内部的其他变异源选择检验员的试验组，例如他们的生产班次、技能水平或类似产品的检验经验。
- b) 被选中进行测试的检验员至少应熟悉产品，否则将成为过于保守的测试。例如，缺乏产品经验可能会导致误报率的增加。
- c) 记录选择不同的检验员组进行测试的文件。

4.2.6.9 步骤 9

准备试验样品：

- a) 征集代表单元

在早期及开发过程中收集有代表性的缺陷设备/样品，为 ATMV 试验做好准备。此时，临界样品尤其值得搜集。但是，请注意，如果行业专家无法就样本的符合或不符合达成一致意见，只会

在试验中造成问题。相反，应选择相对于验收标准能够代表“刚刚通过”和“刚刚失败”的样本。

- b) 对人造缺陷样品是否充分代表密封过程、模拟运输，或其他制造过程中自然发生的缺陷进行最佳判断。如果缺陷无法充分复制和/或发生率太低，无法为测试提供样本，则可以根据测试的原理省略该缺陷类型。
- c) 从主计划中估计试验所需的缺陷数量，并尝试获得 1.5 倍于试验所需估计数量。这将允许剔除破损的样品和不太理想的样品。
- d) 样品可能不需要可追溯性。对样品的唯一要求是能准确地描述符合性或预期的不符合性。然而，如果确认方法有困难，或者希望跟踪特定不符合项的输出，则捕捉可追溯性信息可能有助于实现调查目的。
- e) 推荐一个以上的行业专家来对试验中每个样品的状态进行确认。培训师或生产主管也可能是过程缺陷类型方面的行业专家。
- f) 选择适合于特定样品的存储方法。可能的选择包括带有独立标签隔间的处理盒，可重新密封的塑料袋和塑料瓶。有关预处理的要求，见标准化的试验方法。
- g) 在每个零件上标记密码代码可成功地隐藏缺陷类型，但这并不是隐藏零件身份的有效方法。如果检验员能够记住样品的识别号和他们在样品上发现的缺陷，则在检验员第二次拿到样品时，试验就受到了影响。如果每个样品只被每个检验员查看一次，那么在样品上设置代码就不是问题。
- h) 录像检验是一些人工目视检验的另一种选择，特别是可能随着时间而改变的缺陷，如裂纹或异物。
- i) 如果产品非常长/大，如导丝、导管、袋、托盘、容器密封系统（罐子&盖子），且有关缺陷仅在产品的特定部分，可以选择将相关部分与样品的其余部分分离。如果长度或精密度的降低是使整个产品难以检验的因素，则应使用整个产品。例如：泄漏试验，包装中的液体可能会影响测试结果。
- j) 对有缺陷的样品进行拍照或录像，以关键词储存以备将来参考。

4.2.6.10 步骤 10

起草方案，方案内容建议包括：

- a) 目的和范围。
- b) 参考正在确认的测试方法文件。
- c) 引用其他有关文件的清单（如适用）。
- d) 用于 ATMV 的设备、仪器、固定装置等种类清单。
- e) ATMV 研究的基本原理，包括：
 - 1) ATMV 的统计方法；
 - 2) 试验方法测量的特性和 ATMV 覆盖的测量范围；
 - 3) 试验样品和原理的描述；
 - 4) 样品数量、操作员数量、试验次数；
 - 5) 数据分析方法，包括用于数据分析的任何历史统计数据（例如，单边规范限值下计算%P/T 的历史平均值）。
- f) ATMV 验收标准。
- g) 确认试验程序（如样品制备、试验环境设置、试验顺序、数据收集方法等）。
- h) 随机化方法
 - 1) 有多种方法使样本的顺序随机化。在所有情况下，将随机顺序存储在另一列中，然后重复并将第二个随机列表附加到第一个存储的列表中，每个样本都将由同一检验员进行第二次检验。
 - 2) 考虑使用 Excel、Minitab 或在线随机数生成器来创建试验的运行顺序。
 - 3) 从容器中抽取编号，直到容器清空，并记录顺序。
- i) 有些公司对每个样品设置时间限制或对试验的总时间限制，使试验更能代表生产环境的快速要求。如果使用，应在方案中注明。

4.2.6.11 步骤 11

执行方案：

- a) 确保方案执行过程中符合预处理要求。
- b) 避免检验员匆忙完成试验的情况。预估每个检验员需要的时间，并在试验开始前提前规划，使检验员有足够的时间去完成他们的部分，或者与检验员沟通，允许他们在试验期间去午餐或休息。
- c) 向检验员解释正在试验的检验步骤。澄清每个样品是否有一个以上的缺陷。但是，一个样品上有多个缺陷可能会在试验期间造成混淆。
- d) 如果第一个检验员不能正确地识别缺陷的存在与否，则由公司/小组决定是否由其余的检验员继续执行方案。完成方案将有助于确定这些问题是否普遍存在，这可能有助于避免不再失败。另一方面，立即终止 ATMV 可以为每个人节省相当多的时间。
- e) 如果在试验过程中出现失败，不宜改变抽样计划⁷⁾。例如，如果最初的 β 错误抽样计划是 $n=45$ ， $a=0$ ，并且发生了一个失败，试验期间将抽样计划更新为 $n=76$ ， $a=1$ 是错误的，因为正在执行的抽样计划实际上是一个双抽样计划， $n_1=45$ ， $a_1=0$ ， $r_1=2$ ， $n_2=31$ ， $a_2=1$ 。这导致 LTPD 变成了 5.8%，而不是原来计划中的 5.0%。
- f) 准备备用样品以防有缺陷的样品损坏。
- g) 所有检验员同时进行试验具有风险，因为管理员将保持跟踪哪个检验员拥有的未标记样品。
- h) 在每个检验员试验之后复查分类错误的样品，以确定检验员是否发现了准备团队遗漏的缺陷。

4.2.6.12 步骤 12

分析试验结果：

- a) 舍弃错误缺陷代码或缺陷类型
以下情况可能会出现：检验员用方案中没有包含的词描述缺陷。确认团队需要确定所使用的名词是否与这个特定缺陷对应的任何列出的名称是同义词。如果不是，那么试验失败。如果名词与缺陷匹配，则在报告的偏离部分记录这个异常。
- b) 在计算性能表现时剔除数据
如果在试验完成后发现缺陷，有两个建议选择。首先，如果有必要，检验员可以稍后对替换部件进行试验。或者，如果个别试验的结果不会改变抽样计划的最终结果，则可以跳过替代试验。这种原理宜在报告的偏离部分加以记录。
 - 1) 例如，考虑一个 $n=160$ ， $a=13$ 的 α 抽样方案，该方案设计为满足 12% 的 α 错误率。在所有检验员完成试验后，确定其中一个合格的样品有缺陷。在对该样品进行的 6 次试验中有 5 次识别出了该缺陷，即 6 次试验中有 1 次称该样品为合格样品。这 6 次试验的结果需要被划出，但如果剩下的 154 个合格的试验几乎没有失败，仍然满足 12% 的 α 错误率，那么就没有必要进行替代试验。同样的原理也适用于 β 错误抽样方案中的缺陷样本。
 - 2) 如果真空衰减试验样品未能通过泄漏试验，在这种情况下，作为方案的一部分，过程可能是将样品送回产生有缺陷样品的公司，以确认是否真有缺陷。如果发现该样品不再具有所期望缺陷类型的代表性，则在计算中舍弃该样品。

4.2.6.13 步骤 13

完成确认报告：

- a) 确认试验通过时
如果 ATMV 很难通过或需要特别的检验员培训，考虑增加评定员能力验证 (PT)，以确保检验员能够胜任该过程的检验。
- b) 确认试验失败时
 - 1) 重复确认
对于 ATMV 的失败次数没有限制。然而，宜使用一些常识，因为大量的尝试似乎是一种“试验直到通过”的方法，并可能成为一个审核标志。因此，一个好的经验法则是在执行

7) 对于某些类型的试验仪器，在执行确认过程中可以验证其性能。执行此操作不会被认为修改了样本量。例如：高压检漏或真空衰减。

前进行空运行或可行性评定，优化评定员培训和试验方法，降低方案失败的风险。如果 ATMV 失败，确认团队的成员可以自己进行试验。如果确认团队通过，那么有些信息没有清楚地传达给检验员，需要进行额外的沟通来识别混淆。如果确认团队也未能通过 ATMV，这就强烈地表明目视检验或属性试验方法还没有准备好发布。

2) 用户错误

——ATMV 试验错误的例子包括：

- 显微镜放大倍数设置错误。
- 在 ATMV 期间，由于样品混淆，样品可追溯性受到影响。

——试验失败表明需要减少检验员之间的变异。关键是理解试验失败的原因，纠正问题并记录基本原理，这样后续的试验就不会出现“试验直到通过”的情况。

——如果正在试验的检验员与先前 ATMV 的相同，则其后的 ATMV 应尽量不要使用相同的样本。

——与发生分类错误的检验员进行面谈，以了解他们的错误是由于对可接受准则的误解还是单纯的疏忽。

——为了提高缺陷检测/试验方法的熟练程度，以下是一些建议的最佳实践：

- 在作业指导书中规定检验员的检验顺序，如从近端移动到远端，或先内后外。
- 当对一个部件或组件的多个位置进行特定属性的检验时，提供一个带有顺序编号的可视模板，以便在检验过程中遵循。
- 将显微镜图像转移到视频屏幕，以便观看。
- 如果在一个检验步骤中需要查找太多的缺陷类型，有些缺陷可能会被遗漏。将任何与过程不相关的检验移回到可能产生缺陷的过程上游。

——当检验员误解准则时，需要更好地区分合格品和不合格品。以下是一些建议：

- 与检验员和培训人员一起评审缺陷的目视标准。
- 确定图表是否比照片提供的信息更多。
- 改变显微镜的放大倍数。
- 如果 ATMV 失败是因临界缺陷被错误地接受，则将制造可接受准则阈值降低到一个更保守的水平。这将潜在地增加 α 错误率，无论如何，它通常允许更高的错误率，但 β 错误率应该会降低。

——考虑使用属性一致性分析来帮助识别 ATMV 失败的根本原因，因为它是评定多个评定员给出的标称或顺序评级是否一致的好工具。该分析将计算每个评定员的重复性和多个评定员之间的再现性，类似于变量试验方法的 gage R&R。

4.2.6.14 步骤 14

确认后活动：

a) 试验方法变更

- 1) 如果要求、标准或试验方法发生变化，则需要评估其他两个因素的影响。
- 2) 例如，许多属性试验方法（如目视检验）对被测试设备的形式、适用性或功能没有影响。因此，很容易忽略的一点是，需要仔细评估这些变化（记录在设计图纸上、视觉设计标准上、视觉过程标准上的试验方法准则的变化）可能对设备性能产生的影响。
- 3) 好的做法是召集操作和设计方面的代表，一起审核提出的变更和考虑变更可能产生的结果。
- 4) 例如，在设计验证构建过程中使用的对初始视觉检验标准的更改，可能无法在完成模拟运输之前识别出缺陷。暴露于与模拟运输过程相关的冲击、振动和热循环可能会加剧在此初始检验过程中遗漏的应力。因此，了解上游使用的视觉标准的变更可能对下游检验产生的影响很重要。

b) 增强试验方法确认——有时一个新的缺陷是在 ATMV 确认之后被识别出来。有多种方法可以确认新失效模式的检测。

- 1) 选择#1—在添加新准则的同时重复整个确认过程。

——优点：最终结果是一个完整的、独立的确认，完全代表了最终的检验配置。

——缺点：该工作量过大，相当于对整个 ATMV 进行重新确认，都是为了增加一个单一的检验准则。此外，如果 ATMV 在一种已存在的失效模式下出现失败，则会对确认过的 ATMV 以及通过该试验方法批准的历史器械产品产生怀疑。

2) 选择#2—运行仅与新准则相关的缺陷的全功效 ATMV。

——优点：由于检测其他缺陷的能力已经得到确认，因此该方法将重点放在新引入的准则上。

——缺点：缺陷类型在试样中的分布宜基于表示历史废料的比。因此，一个只包含一个缺陷的 ATMV 将完全超过原 ATMV 中其他缺陷代码的样本量。其次，如果可用的检验员数量有限，要找到足够的不合格样品很快就会成为一项繁重的工作。

注：考虑对缺陷类型进行基于风险的分析，以确定适当的样本量。

3) 选择#3—在较小的样本量下运行一个只有新缺陷的增强 ATMV。

——优点：这种方法结合了前两种方法的优点而非缺点。重点是在新的检验准则下的缺陷，但在充分运用检验员知识的样本量下。增强后的报告可以指向原来的 ATMV 和更新的 MVR，这样未来的 ATMV 就可以知道新的缺陷，将其包含在未来的报告中。

——缺点：有人可能会说检验员没有同时对新旧缺陷的混合进行试验。然而，为了评估 α 和 β 的错误率，检验员仍然会面临合格和不合格部件的混合。因此，增强方法仍然挑战检验员对目视要求的理解。

4) 选择#4—基于随着时间推移新获得的过程知识，运行一个增强的 ATMV。

——优点：可能会减少样本量，因为随着时间的推移，已经收集了更多的数据，可以更好地了解缺陷的真实发生率。

——缺点：可能会增加样本量，有更多关于缺陷发生率的数据。

——利用现有的属性试验方法。

在适当的时候利用属性试验方法是一种聪明、有效的方法。特别是，当被检验的组件与之前确认过的组件完全相同或足够相似时，宜考虑利用。当组件被用在下一代器械或当产品被转移到另一个地点时，可能会发生这种情况。

——利用 ATMV 的建议要求。

- 使用相同的工具、检验准则和培训方法。例如，显微镜应该设置在相同的放大倍数。如果放宽或加严新器械的检验，则应重复 ATMV。
- 新产品仍需满足最低性能要求。因此，如果旧产品上的缺陷被评级为低风险级，但新产品上的相同缺陷被认为是中风险或高风险，那么不能使用现有的 ATMV，除非原有的 ATMV 是按照高风险需求级别实施的。
- 检验过程或可接受准则不能有任何变更或“改进”。一个人的澄清是另一个人的混淆。
- 利用 ATMV 时不一定需要方案，只需要一份报告。这里的信息是，宜有一种统一的记录方式，表明该试验方法已经被引用，但是这个活动不需要触及方案。相反，只需更新原始的试验方法报告，说明新增的产品、生产线或生产地点也在使用此已完成确认的试验方法。

4.3 变量试验方法确认 (VTMV)

4.3.1 变量试验方法确认的目的

在假定试验结果准确和精确之前，宜对变量试验方法进行确认。试验方法的变异存在于所有试验中，应评估其对试验结果和/或产品规范公差的影响。进行变量试验方法确认的目的是：

- a) 证明试验方法的精密度足以满足所测量的试验特性。
- b) 提供试验设备运行一致和试验程序性能一致的客观证据。

VTMV 涵盖了使用量具或试验设备的精确度要求。尽管本文件中没有明确规定量具或试验设备的准确性，但应按照各公司或设备制造商推荐的校准程序和设备鉴定过程对其准确性进行评估。在校准和设备鉴定完成的前提下，一个 VTMV 的成功结果将确保试验方法(包括试验设备)做出准确和精确的测量。

4.3.2 试验方法变异

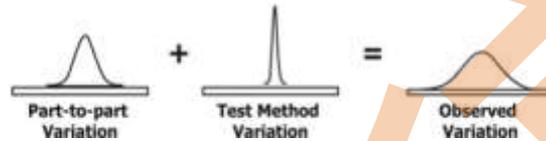
一般情况下，观察到的试验结果包括两种来源的变异：零件间变异和试验方法变异。观察到的零件间和试验方法的变异之间的关系可以用数学方式描述：

$$\sigma_{Part}^2 + \sigma_{TM}^2 = \sigma_{Observed}^2 \dots\dots\dots (6)$$

式中：

- σ_{Part} —— σ 零件间变异；
- σ_{TM} —— σ 方法变异；
- $\sigma_{Observed}$ —— σ 观察变异。

或者，如图1所示。



标引序号说明：

- Part-to-part Variation —— 零件间变异；
- Test Method Variation —— 方法变异；
- Observed Variation —— 观察变异。

图1 观察变异

试验方法变异 σ_{TM}^2 由两部分组成：重复性和再现性。重复性是指同一评定员在相同的试验条件下多次重复测量同一部件产生的波动，重复性波动主要反映量具本身的变异。再现性是指不同的评定员（或自动化试验时的不同的试验设备）测量相同部件的变异性。

在Gage R&R研究中，采用方差分析（ANOVA）或广义线性模型（GLM）将观察到的检验结果的变异分为真实零件之间的变异和试验方法的变异（见图2）。

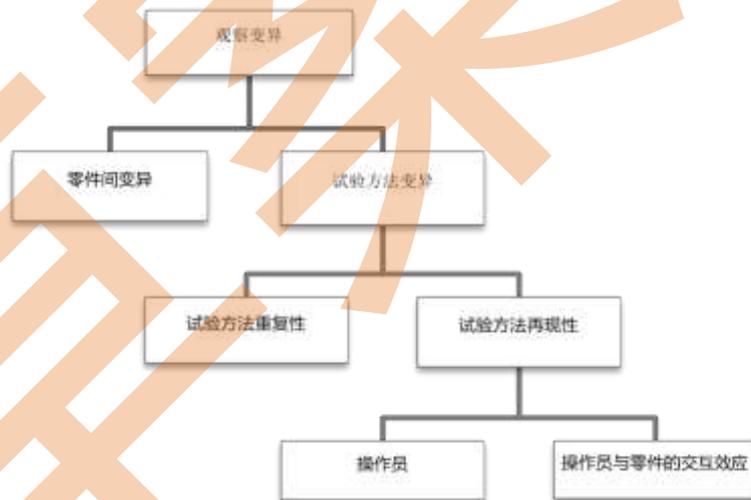


图2 观察变异

4.3.3 变量试验方法确认步骤

4.3.3.1 步骤 1

准备试验方法文件：

- a) 在执行VTMV方案前，需要发布书面试验方法文件或书面试验方法作业指导书。
- b) 当对已在最终版本文件中发布的现有试验方法进行确认或再确认时，需在参考VTMV方案和/或报告明确版本。

4.3.3.2 步骤 2

设备鉴定：

- a) 按照公司的程序完成 VTMV 所用的所有设备、仪器、夹具等的鉴定。
- b) 如果试验设备中存在软件，考虑软件是否需要按照适当的法规进行确认。

4.3.3.3 步骤 3

选择 VTMV 方法和可接受准则（见表 2）：

表 2 选择 VTMV 方法和可接受准则操作注释

行动	注释
选择 VTMV 合适的统计方法	三种 VTMV 常用的统计方法如下： ——能力研究； ——Gage R&R 研究； ——功效/样本量研究（仅用于确认比较检验方法）。 除了这三种方法，VTMV 还可以通过利用现有的 TMV 来完成。
建立 TMV 验收标准	参考现有的 TMV 可接受准则，或考虑使用行业公认的标准。
制定一个 VTMV 研究	见下面四节，了解如何使用选定的统计方法策划一个 VTMV 研究。

a) 能力研究

- 1) 在缺乏统计学依据的样本量情况下，要从两位及以上评定员那里收集至少 $n=30$ 个测量数据点。
- 2) 如果现有数据包含至少 30 个样本，并且在这种情况下不需要 VTMV 方案，则可以使用现有数据。但是，应在 VTMV 报告中纳入原始数据来源的参考。
- 3) 如果一个试验方法是由一个指定的评定员执行，则允许仅使用来自该评定员的数据，以满足 $n=30$ 的样本量要求。但是，试验方法文件和试验方法确认报告应该清楚地说明这个限制。为了在之后对该试验方法新增受过培训的评定员，需要对原评定员的数据进行重新评估，然后修订试验方法文件和试验方法确认报告。

b) Gage R&R 研究

- 1) Gage R&R 研究是一种统计方法，用于：
 - 估计试验方法的重复性、再现性和总变异性；
 - 评估相对于被测量零件或产品的公差范围，或相对于整个工艺变异或整个研究变异，试验方法的精度是否足够。
- 2) Gage R&R 研究的样本量
 - Gage R&R 研究的目的是估计 $\sigma_{\text{重复性}}$ 和 $\sigma_{\text{再现性}}$ 。样本量确实影响标准差（或方差）估计：

样本量	影响
部件的数量	$\sigma_{\text{重复性}}$ 和 $\sigma_{\text{再现性}}$
操作员的数量	$\sigma_{\text{再现性}}$
试验的次数	$\sigma_{\text{重复性}}$

——一般来说，样本量越大，标准差估计越准确。这是因为样本标准差统计量遵循与样本大小相关的自由度卡方分布（ χ^2 分布）。更大的样本量（更高的自由度）将导致卡方分布范围更小，因此标准差的置信区间更窄。

——《Gage R&R 设计与分析》一书中描述的 Gage R&R 样品量的详细数学评估⁸⁾使我们得出以下结论：（部件数量）×（操作员数量）×（试验的次数 - 1） ≥ 15 。

——在 Gage R&R 研究中，建议 3 位及以下的操作员，以降低对再现性夸大估计的风险，再现性具有更严格的阈值。

3) 非破坏性试验方法

8) Richard K. Burdick, Connie M. Borror 和 Douglas C. Montgomery. Gage R&R 研究的设计与分析.2005.工业和应用数学学会，费城，宾夕法尼亚州。

——在 Gage R&R 研究中使用至少 3 个试验样品。如果 VTMV 的范围包括多个产品/型号/标签，可以使用以下三种方法之一来选择样品进行试验：

- 使用对 VTMV 可接受准则提出最大挑战的产品/型号/标签的样品（例如，具有最高试验方法变异性但最严公差范围的产品）。
- 使用插入法技术，运行两个或两个以上的 Gage R&R 研究，每一个针对代表可能的极端条件的产品/型号/标签。使用该策略对于充分确保试验方法在整个极端条件下的性能是必要的。
- 将不同产品/型号/标签的样品集中在一个 Gage R&R 研究中。这种方法适用于合并的产品/型号/标签具有相同的公差，或者可以假定不同产品/型号/标签之间的试验方法变异具有可比性。

——从所有合格的试验方法评定员中以随机或涵盖广泛的经验水平的方式，选择至少 2 名评定员。

——按照表 3 确定每位评定员和每个试样的最小试验次数（重复读数）。

——随机化试验顺序，如有必要，准备数据表。

4) 破坏性试验方法

——对于破坏性试验方法，很难将试验方法的重复性与零件之间的变异分开。但是，在破坏性 Gage R&R 研究中仍然可以评估试验方法的重复性和再现性。可以使用以下四种方法：

- 使用替代品。替代品是代表预期使用的试验方法的实际产品/部件的替代品。替代品可以为非破坏性试验进行重复测量，或者替代品的零件间的变异与测量误差的可重复性部分相比可能足够小，对总测量误差的影响很小。要使用替代品，VTMV 方案中应包括替代品的描述及其对试验方法预期用途代表性的证明。替代品数量、评定员数量和试验次数应满足表 3 的要求。
- 使用主组。主组（也称为主批、主批次）是具有较小零件间变异的均质部件的集合（例如，从单个生产批次中提取的部件）。在 Gage R&R 研究中，每个主组将被视为假设的“试验部分”，主组中的每个单元将被视为一次试验。注意，在这个方法下，估计的可重复性不仅包括真实的试验方法重复性，还包括主组内部件间的差异。为了在 Gage R&R 研究中使用主组方法，应该在 VTMV 方案中对主组进行明确定义和记录。主组数、评定员数量和试验次数见表 3。

表 3 Gage R&R 研究建议的最小试验次数

#评定员	#测试样品编号					
	3	4	5	6	7	≥8
2	5	4	4	3	3	2
3	4	3	3	2	2	2
≥4	3	2	2	2	2	2

- 如果主组同质性不能被假定，仍然可以使用主组方法进行 Gage R&R 分析来估计试验方法的再现性。然而，通过在主组中包含大量的部件之间的变异，试验方法的重复性将被过度估计。在这种情况下，可以使用替代品、标准品或其他技术进行一个独立的研究，以估计试验方法的重复性。检验方法的总标准差将根据平方和开方法（RSS）计算⁹⁾：

$$\sigma_{TM} = \sqrt{\sigma_{\text{repeatability}}^2 + \sigma_{\text{reproducibility}}^2} \dots\dots\dots (7)$$

式中：

- $\sigma_{\text{repeatability}}$ —— σ 重复性
- $\sigma_{\text{reproducibility}}$ —— σ 再现性

9) Draper, N.R.; Smith, H. (1998)。应用回归分析（第三版）。John Wiley。

- 在某些情况下，当试验样本的测量不能完全重复时，可使用工程或统计模型来描述随后测量的变化。这些模型可以帮助估计真正的试验方法的可重复性。使用这种方法通常涉及先进的科学或统计方法，宜向行业专家咨询以获得援助和批准。

注：对于常用的破坏性包装试验方法，如YY/T 0681.2-2021密封强度（热密封剥离）试验，重要的是要了解，当在拉力测试仪（本试验方法使用的仪器）上进行试验方法确认时，应该关注方法的变异而不是材料。因此，如果可能的话，尽量使用变异最小的材料，使其仍能覆盖要试验的拉力范围。

c) 功效/样本量研究

功效/样本量研究仅适用于确认比较检验方法，该方法用于使用统计检验比较总体均值（例如，单样本t检验、双样本t检验或方差分析）。运行功效/样本量研究的目的是确保试验方法的可变性不会掩盖均值（“信号”）的差异，该均值（“信号”）应由足够功效（ $\geq 80\%$ ）的统计试验检测到¹⁰⁾。

进行功效/样本量研究：

- 1) 建立用于比较均值的实际显著性水平 Δ ，并在VTMV方案和/或报告中记录 Δ 的基本原理。
- 2) 根据最低 $n=15$ 个测量值估计总标准差。这些测量值应该从5个及以上的独立样本中收集。
- 3) 计算需要的样本量 n 进行统计检验，以检测 Δ ，其功效 $\geq 80\%$ 。
- 4) 如果样本量 n 可接受，则接受试验方法。

d) 使用现有数据

当满足以下要求时，VTMV可以使用已有数据：

- 1) 现有数据应以可追溯其原始来源的可控方式储存（例如，储存在设计历史文档（DHF）中的数据；经批准的电子实验室笔记本（ELN）；或作为经批准的变更控制文件的附件）。
- 2) 现有数据应从测量产品或部件中收集，这些产品或部件在试验方法的确认范围内，或具有试验方法预期用途的代表性。
- 3) VTMV报告可包括现有数据如何满足这些要求的理由。

4.3.3.4 步骤4

完成VTMV方案：

- a) 除使用“仅报告”方法外，按照GB/T 19633.1要求，在执行确认试验之前需要发布VTMV方案。VTMV方案至少可以包括以下内容：

- 1) VTMV方案的目的和范围。
- 2) 参考正在确认的测试方法文件。
- 3) 引用其他有关文件的清单（如适用）。
- 4) 用于VTMV的设备、仪器、固定装置等的种类清单。
- 5) VTMV研究的基本原理，包括：
 - VTMV的统计方法；
 - 试验方法测量的特性和VTMV覆盖的测量范围；
 - 试验样品的描述及原理；
 - 样本数量、评定员数量、试验次数；
 - 数据分析方法，包括用于数据分析的任何历史统计数据（例如，在单边规范限值下用于计算%P/T的历史平均值）。
- 6) VTMV可接受准则。
- 7) 确认试验程序（如样品制备、试验环境设置、试验顺序、数据收集方法等）。

4.3.3.5 步骤5

评定员准备：

- a) 从所有有资质的试验方法评定员库中选择所需数量的评定员，其涵盖广泛的技能和经验水平。

10) Bausell, R. B. and Li, Y (2002)。实验研究的功效分析：生物、医学和社会科学实践指南。剑桥大学出版社，纽约。

- b) 在执行 VTMV 方案前对评定员进行试验方法培训。培训应包括任何试验环境的特殊设置和 VTMV 专属设备的操作，这些在运行试验方法之前将由评定员负责。培训应根据公司的培训要求进行记录。

4.3.3.6 步骤 6

准备试验样品。收集所需数量的试验样品，准备进行确认试验。

- a) 除使用代表性替代品或标准品外，供试品应在 VTMV 范围内。
- b) 对于使用能力研究或功效/样本量研究的 VTMV，试验样本应来自代表实际设计和/或生产过程的正常制造。
- c) 对于使用 Gage R&R 研究的 VTMV，试验样品可能包括标称、边界和超规格单元。以便在跨越公差范围的情况下评估试验方法的性能。
- d) 以一致的方式准备试验样品。
- e) 仔细标记和/或控制每个试验样品，以防止混淆和偏差。

4.3.3.7 步骤 7

执行 VTMV 方案：

- a) 完成设备和试验环境的设置。
- b) 检查量具的校准和/或确认（如适用）。
- c) 按照 VTMV 方案规定的试验顺序进行。
- d) 同时记录试验结果。
- e) 保持每项测量结果可追溯至评定员和试验样品。
- f) 记录任何与方案的偏差，以供审查。
- g) 尽可能保留试验样品和其他相关信息，直至完成确认报告并获得批准。
- h) 如果在确认试验过程中发现需要对试验环境和设备/夹具设置进行任何调整，应评估该调整对后续测量的影响。

4.3.3.8 步骤 8

分析 VTMV 结果：

- a) VTMV 结果分析宜遵循行业公认的统计方法。应评估任何 VTMV 方案偏差对 VTMV 结果的影响。
- b) 能力研究
利用适当的分布和/或转换评估分布拟合性并计算能力指数（单边规格限的 Ppk 或 Cpk，双边规格限的 Pp 或 Cp）。
- c) Gage R&R 研究
 - 1) 进行 Gage R&R 研究（交叉）分析，进行非破坏性 Gage R&R 研究。
 - 2) 进行 Gage R&R 研究（嵌套）分析，进行破坏性 Gage R&R 研究。
 - 3) 如果 Gage R&R 研究涵盖不同的产品/型号/标签，使用最坏情况公差范围或最坏情况工艺标准偏差来计算 %P/T 或 %R&R。
 - 4) 如果试验输出只有单边规格限，则 %P/T 计算需要使用试验输出的历史平均值或 Gage R&R 研究平均值（见上述的 %P/T 定义）。如果使用历史平均值，则宜根据有效的数据源（例如，来自自己发布的技术报告、设计验证报告或过程确认报告的数据）进行计算。Gage R&R 研究平均值仅在代表试验输出的标称平均值时使用。
- d) 功效/样本量研究
 - 1) 功效和样本量计算采用 95% 置信水平。在 VTMV 方案和/或报告中，可以根据适当的理由使用不同的置信度。建议将置信度与故障相关的风险联系起来。例如，如果评估一个对产品功能至关重要的特性，那么就使用更高的置信度，比如 95%。
 - 2) 在计算功效和样本量时，如果适用，采用单边备选假设。
- e) 完成 VTMV 报告——如果满足了 VTMV 可接受准则中所定义的最低性能要求，接受结果并完成一份 VTMV 报告，其中包括：

- 1) 范围部分，明确标识 VTMV 涵盖的产品/部件和/或测量范围；
- 2) 参考试验方法文件和其他相关文件；
- 3) 用于 VTMV 的设备、仪器、固定装置等清单；
- 4) 任何与 VTMV 方案有适当理由的偏差和对确认结果的影响的评估；
- 5) 与 VTMV 可接受准则中有关的试验结果的统计分析（图形和/或数字输出）；
- 6) VTMV 结论；
- 7) 将原始数据附件添加到 VTMV 报告的附录页，或根据现场实践，将原始数据纳入经批准的电子实验室笔记本（ELN）或数据存储系统中；
- 8) 如果统计分析中排除了一些原始数据，应包括适当的理由，解释为什么不包括这些数据和为什么不需要重复确认（例如，条件是什么，为什么会发生等等）；
- 9) 如果 VTMV 出现失败，包括对失败的调查和所采取的纠正措施；
- 10) 在确认试验之后和 VTMV 报告发布之前对试验方法所作的任何更改的说明。对于不需要再确认试验的变更，应提供不需要再确认试验的理由。

附录 A
(资料性)
可接受准则示例

A.1 下表提供了一个示例，说明如何处理与正在确认的测试方法的可接受准则相关的风险。置信度、可靠度和风险级别应根据具体情况来确定，这里仅作为示例列出。

表A.1 可接受准则示例

数据类型	置信限	风险指数水平	最低水平
			主要方法
			LTPD /可靠度
β 错误	90% (第十百分位)	高	1%/95%
		中	10%/90%
		低	3%/97%
α 错误	用户定义	0	团队根据基本原理决定
	90% (第十百分位)	高、中、低	用户定义的最小样本量限制
	用户定义	0	团队根据基本原理决定

附 录 B
(资料性)
ATMV 工作表示例

B.1 这个示例不是强制性的。

表B.1 ATMV 工作表示例

项目名称		测试日期	
项目编号		运行#	
按照工序的检验#			
使用到的工具		设备#	校准到期时间
检验员姓名及签名			
样品#	结果处置 (含原因的通过或失败的结果)	样品#	结果处置 (含原因的通过或失败的结果)
1		15	
2		16	
3		17	
4		18	
5		19	
6		20	
7		21	
8		22	
9		23	
10		24	
11		25	
12		26	
13		27	
14			

参 考 文 献

- [1] ASTM F3263-17 Standard Guide for Packaging Test Method Validation
- [2] ASTM E691-22 Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method
- [3] ASTM E177-20 Practice for Use of the Terms Precision and Bias in ASTM Test Methods
-

