ICS 67.240
CCS X 04

中华人民共和国国家标准

GB/T XXXX—XXXX/ISO 20784:2021

感官分析　感官与消费品宣称证实导则

Sensory analysis—Guidance on substantiation for sensory and consumer product claims

（ISO 20784:2021,IDT）

（征求意见稿）

XXXX - XX - XX 发布　　XXXX - XX - XX 实施

国家市场监督管理总局
国家标准化管理委员会　发布

目　　次

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。

本文件等同采用了 ISO 20784:2021《感官分析　感官与消费品宣称证实导则》，一致性程度为等效。

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。

本文件由全国感官分析标准化技术委员会（SAC/TC566）提出并归口。

本文件起草单位：

本文件主要起草人：

感官分析　感官与消费品宣称证实导则

1　范围

本文件给出了对食品和非食品产品及产品包装上、以宣传为目的的感官宣称进行证实的导则。

本文件给出了感官宣称与其他类型宣称的区别，提供了感官宣称的分类和不同类型感官宣称的示例，以及开展感官宣称证实测试相关的内容，包括示例和参考资料。

本文件适用于产品的感官宣称证实。

本文件不适用于下列情况。

——用于感官宣称证实的不同测试方法的具体或详细要求。

——产品原产地、组分、加工和营养成分相关的事实宣称。

——产品技术特征的事实宣称。

——人类食用或使用产品时与健康、医学或治疗功效、生理功效、结构或功能益处相关的宣称。

——应用仪器对产品进行特性或性能评价的宣称（仪器评价是指不使用评价员或受试者，而使用仪器对产品的特性或性能进行评价）。

——服务类项目相关的宣称（例如房屋清洁服务、航空服务、汽车服务等）。

——关于大件/慢速消费品相关的宣称（例如汽车、冰箱、灶具等）。

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 10221—2021　感官分析　术语（ISO 5492: 2008，IDT）

3　术语和定义

ISO 5492 界定的以及下列术语和定义适用于本文件。

ISO 和 IEC 维护的用于标准化的术语数据库网址如下：

——ISO在线浏览平台：https://www.iso.org/obp

——IEC 电子百科：http://www.electropedia.org/

3.1

感官宣称　sensory claim

向消费者宣传一个产品的感官特性（例如“具有焙烤风味”）、功能特性（例如“去油”）/性能特性（例如“口气清新更持久”），以及消费者对该产品的情感响应（例如“消费者偏爱品牌 X”）和使用产品时（包括使用前、使用期间或使用后）的感知响应（例如“皮肤看起来更年轻”）的表述。

注：感官宣称包括在公共场所发布的任何形式的广告信息。广告信息包括产品包装、印刷品或数字媒体（如电子、电视或视频等形式）等多种形式。目的是将该产品的感官特性告知给产品的潜在用户或购买者，或强调消费者在使用产品（包括使用前、使用期间或使用后）时的感知。这种广告信息是让潜在用户或购买者了解产品特性，进而产生购买、消费或使用的欲望。

3.2

情感宣称 affective claim

用户或潜在用户在使用产品（包括使用前、使用期间或使用后）时与其喜好或情绪响应相关的表述。

注：该响应包括消费者在使用产品（包括产品使用前、使用期间或使用后）时，由产品引发的消费者喜好、态度、认知或情绪上的响应。最常用的喜好响应测试是喜好测试或偏好测试。态度响应通常是指消费者未来更愿意购买该产品，或者是消费者认同该产品具有某种突出特性或能提供特定情感体验的声称。

3.3

感知/性能宣称 perception/performance claim

描述产品引起的某种感知特性或者预期效果的相关表述。

示例 XXX 产品粘稠（感知）且无残留（性能）。

3.4

鼓吹 puffery

一种极其宽泛、模糊和主观的陈述，由于过度夸大导致可信度不高，同时从测量操作或实践角度来看不具有可测性。

3.5

等效性宣称 equivalence claim

声称两个或多个产品在某个或多个产品特性上等同时所提供的表述。

3.6

卓越性宣称 unsurpassed claim

声称产品在一个或多个特性上不会被其他产品超越时所提供的表述。

3.7

优越性宣称 superiority claim

一种比较性宣称，声称产品与另一个或多个产品相比，性能或特性水平更高、喜好度更高或更被偏爱的表述。

3.8

风险 risk

对伤害发生概率和严重程度的综合评价。

注：感官研究人员和利益相关方宜考虑基于感官测试得出的宣称的相关风险。风险是指提出宣称后出现负面后果的概率或可能性。这些负面后果形式并不明确，包括消费者在社交媒体上发布负面帖子或在感官、营销或法律社区内发表评论，来自竞争对手的挑战以及自我监督、监管机构或政府机构等采取的相关行动。在公开提出宣称之前，应识别、讨论和理解风险。

[来源：《ISO/IEC指南》51:2014, 3.9，有修改]

3.9

感官分析方法 sensory analysis methods

一套广泛使用的、有科学依据的感官分析方法，包括差别检验、描述性分析和性能评价方法等。

注：内部有效性和实验室体系控制是感官测试的质量保证，尤其当测试目标是产品特性时更要特别关注和控制。对于同时需要考虑产品和参与者的感官测试，屏蔽产品的品牌标识并进行独立评价是最适宜的方法。测试方法的灵敏度、效应量以及评价员的数量和类型均是感官分析方法中需要考虑的因素（参见 IS0 6658）。

3.10

消费者测试方法 consumer methods

从事产品感官和消费者测试的大多数专业人员所使用的消费者定量测试系列方法，包括情感测试和感知/性能测试。

3.11

代表性消费者样本 representative sample of consumers

测试中使用的消费者群体，通常是从更大的总群中抽取的一个较小群体。该群体的测试结果涵盖了采用总群体测试时可获得的响应变化范围。

注：消费者情感响应测试中，消费者代表性样本特征如下。a) 消费者数量足够多，以保证测试结果涵盖了总群体的情感响应变化范围。b) 使用合格的消费者，即选择产品的真正使用者或消费者、购买者或者接受产品概念的消费者。c) 消费者样本抽取方案的变量要包括人口统计、地理、行为或心理等影响群体差异的相关因素。

3.12

代表性产品样本 representative sample of products

从市场可获取的产品中抽取的用于测试的系列产品，用于产品感官特性、性能特性或产品引起的喜好响应等方面的测试。

注：通常建议研究人员待产品从工厂生产端流通到常用销售端后，再通过零售渠道获取产品后进行测试。样品选择原则是测试产品能代表消费者在市场上所能购买得到的产品。选择足够多数量、批次和不同厂商的产品，保证能够覆盖此类产品的所有变化范围。可以使用新品推广、销售和上市前的样板产品来支持宣称。如若将新样品测试结果用于宣称证实时，需提供新样品与市场产品等同的相关信息。此外，选择产品时，测试产品数量和包括变量的数量取决于广告想要表达的程度，即是否适用于消费者可能会使用到的所有产品。

3.13

客观测试结果 objective test result

采用科学领域中被广泛采用的实验方法开展测试所获得的结果，该测试结果并不依赖于实验者的期望或干预（具有可验证性）。

注 ：客观研究中，数据收集并不会受到实验者的干预，实验设计也允许得到1种以上的可能结果。同样，客观研究中的受访者也不会知晓任何可能影响其对潜在研究目标响应的信息。因此，客观测试结果不受测试人员或测试管理人员的意见或期望的影响，由包括相关变量、并遵循最佳操作实践开展的研究得出，并非预期结论。通常，客观研究结果通过不同研究获得，并且能通过其他相关研究或组合测试结果进行验证。

4 一般要求

感官宣称证实时，宜考虑到下列因素。

a) 感官宣称是基于受试者对产品直接体验的响应记录。

b) 产品宣称旨在向潜在购买者介绍产品特性，阐明产品与竞品的差异并影响购买者的购买决策。

c) 很多国家关于儿童或儿童产品的宣称都受到了严格的监管和限制。

d) 支持宣称的证据是否采用了正确的科学方法以及支持数据的权重和相关性。

注 1：公司的研究人员通常会反复多次测试过自己公司的产品，频繁测试产品时可能会出现某次测试结果与其他测试结果冲突的情况。这种情况下，如果大多数测试结果的分布非常集中时，对宣称的支撑更强。如果单次测试结果在前期获得的结果范围内，也可将其作为宣称的基础。如果没有前期测试结果，与单次结果一致的其他技术信息或证据也能加强对宣称的支持。

注 2：供应商或测试机构开展测试后，如果没有前期测试结果记录导致无法进行比较，提出宣称的公司有责任确定基于单次测试结果得出宣称的风险。

e) 感官宣称必须是基于以下内容的标准化、科学化测量。

——采用感官分析方法确定的产品特性/性能。

——用户对产品特性/性能的喜好、偏爱、感受、态度或感知。

f) 赫尔辛基宣言中给出了需要进行临床研究的宣称定义，即临床试验是“对人体受试者（病人和健康志愿者）所使用的药物或医疗设备进行的系统评价，以证实或揭示试验药物的功效、不良反应以及试验药物的吸收、分布、代谢和排泄，目的是确定试验药物的疗效和安全性”[5]。因此，需要进行临床研究的宣称并不属于感官宣称，不在本文件的讨论范围内。

g) 与关注产品疗效或者产品可能对人体结构或操作产生的潜在影响的研究相比，感官研究主要涉及的是对感官效果的评价，例如特定的口感、香气或外观。

示例：如护肤产品，若研究目的是证明皮肤外观发生变化或改善，可采用裸眼检测和评价员评价，这属于1项感官研究。在没有仪器的情况下，由外部评价员对皮肤触觉特性变化的检测也属于感官研究。

如果护肤产品宣称是改变了皮肤的深层组织/功能或者是整个皮肤表层、真皮层的状况时，这属于1项临床研究而并非感官研究。

注 3：上述2种研究类型存在细小差别。对于化妆品和个人护理用品，“感官研究”和“临床研究”类型区分一定会随着国别和公司而有所差异。本文件中，认为能“改善人体的深层组织/功能”的宣称表述是需要临床研究，因此不在本文件的讨论范围内。

h) “鼓吹宣称”或“夸大宣传”（国家法规中的定义）通常使用非常含糊（例如“这款香水会带你飞”）或者过于夸张（例如“世界上最舒适的鞋子”）的表述，没有人会当真。由于表述含糊不清或者出于实用考虑，鼓吹宣称无法使用科学测试数据支撑。

i) 计划开展感官宣称测试的研究人员必须清楚了解感官宣称相关的监管机构、管理法规、媒体或潜在竞争对手对感官宣称可能的反应。研究人员宜在进行研究方案设计和分析时就提出对相关法规问题和竞争对手可能反应等问题的解决措施。

5 指导原则

5.1 总则

使用感官或消费者测试方法开展感官宣称证实时，调查人员应考虑 5.2-5.10 中给出的9项原则，以获得足够可靠的测试结果来证实宣称内容。

5.2 查阅政府的法律法规

开展宣称测试前，应先查询拟发布宣称所在国（地区）/的政府、监管机构、广告业和媒体对感官宣称的要求和标准。

5.3 阐述宣称主要内容并设计测试方案

设计支撑宣称主要内容的研究方案。建议测试前先确定宣称的表述内容，以便合理设计实验。依据预期的宣称内容，选择主要的研究目标（例如描述特性或是消费者响应）。理想情况下，宜限制目标的数量，避免产生多重效应[2]。注意，研究中评价项目越多，出现虚假或矛盾结果的可能性就越大。检验统计量增加与概率变化的函数关系参见附录 B。

注 在有些国家，研究之前先确定感官宣称的类型是十分必要。

5.4 确定宣称类型：单个产品测试或比较性测试

确定宣称类型时宜先确定测试是使用单个产品、成对产品还是多种产品。非比较性宣称宜采用一元测试，如果是产品类别宣称，宜使用成对产品或一个产品与多个产品的比较性测试。

注：比较性宣称时，要求研究人员明确知道用于比较的产品数量和类型。比较性宣称测试中，对于竞品选择的具体指导，参见 ASTM E1958。

5.5 确定决策标准

决策标准宜提前进行界定，且清楚明确。

5.6 确定相关产品集

对于比较性宣称，宜提前确定好相关的产品集。

5.7 确定相关消费者或评价员群体

对于消费者测试，宜在研究开展前先确定相关用户、当前用户、购买者或潜在消费者群体。对于以产品为重点的测试，提前确定经过训练的、具有相关评价资质的评价员群体。

5.8 确定证据强有力的程度

证据宜足够强有力以应对预期中的挑战。

注 研究人员知晓到宣称的证据需多强有力，以及宣称可能受到政府、监管部门或者竞争对手质疑的风险。所有宣称在确定证据强有力程度时要考虑下列因素。

a) 相同测试结果重复出现多少次时，足以让宣称方更加确定结论是“可靠的”。

b) 支持宣称证据的可信程度，例如确认该结果是否与其他来源的信息（结果）一致。

c) 受访者群体和产品样本是否具有足够的相关性和代表性。

d) 研究中是否很好的管理了实验变量和混杂变量。

e) 测试时遵循上述最佳实践原则的程度，包括测试前选择和遵循的主要目标（见 5.3）。

5.9 保证公正性

提出宣称的组织方宜确保测试方法、测试条件和测试执行的公正性。例如，两种同时生产的产品以相同方式制备和提供，呈送时遵循平衡原则，确保两种产品均采用了相同的测试程序。

5.10 保证可靠性

宜遵循最佳实践原则来确保研究的客观性和公正性，研究结果重复性好，且足够可靠。

6 感官宣称分类

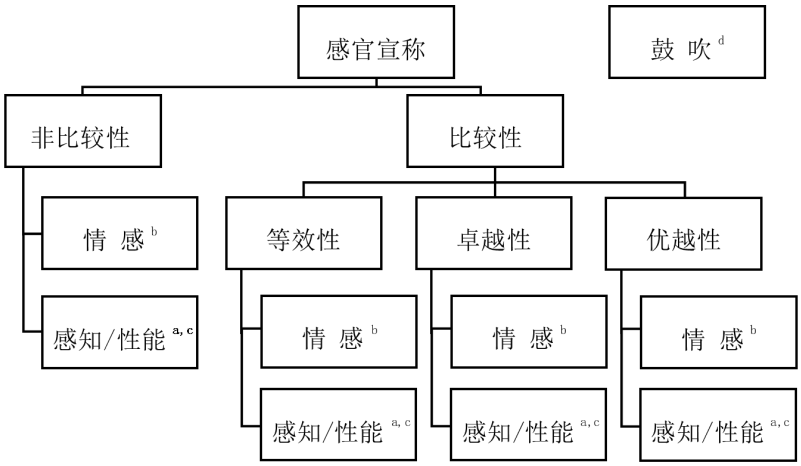
6.1 分类

感官宣称分类有利于形成清楚合理的证据以及选择适宜的测试方法。感官宣称分类如图 1 所示，不同宣称类别的主要差异在于下列三种情况。

a) 是单个产品的宣称还是与竞品的比较宣称。

b) 是消费者对产品或性能的情感响应宣称还是感知响应宣称。

c) 受访者是消费者还是经过训练的评价员。



a 使用经过训练的评价员或相关的消费者开展测试。

b 由相关的消费者（当前购买者或潜在用户）参与情感宣称测试。

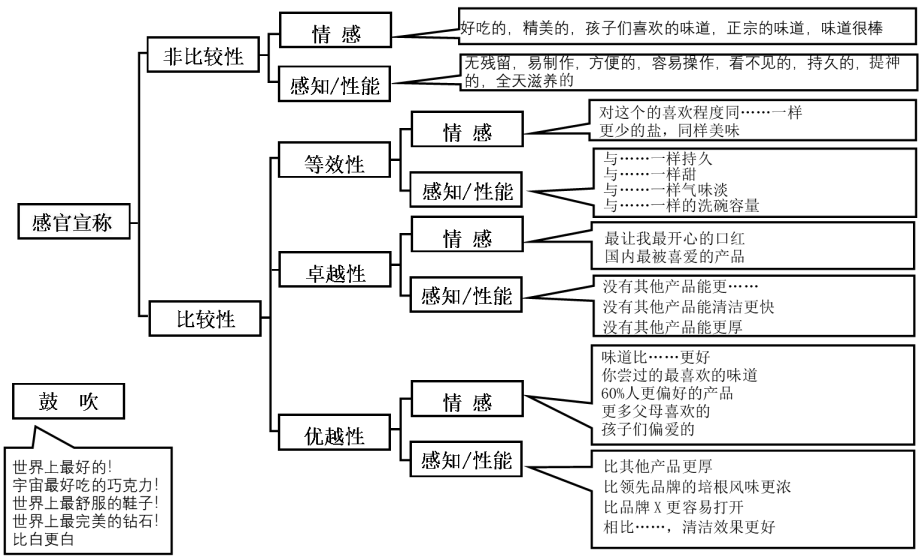
c 等效性、卓越性和优越性宣称中使用的受访者类型取决于是情感宣称还是感知/性能宣称。

D 鼓吹是一种特殊类型的宣称，极度夸大或含糊不清，无法通过感官方法进行测试。

图 1 感官宣称的分类

6.2 表述

图 2 给出了不同类型感官宣称的表述示例，图中仅列出了一小部分示例。宣称的表述会因产品类别、期望的目标消费者和感官宣称实施所在国别而有所差异，主要取决于感官测试是如何直接、明确地支持宣称的表述。



注：感官宣称类型很多，表述方式也非常多样。图2列出了不同类型感官宣称的表述示例。关于感官宣称证实的示例，参见附录A。

图 2　不同类型宣称表述的具体示例

6.3　非比较性感官宣称

非比较性感官宣称测试包括一元产品或单个产品的测试，有明确的测试目标，依据决策标准具体展开，并使用经过统计手段分析的定量数据。对于情感型的非比较性宣称，采用消费者定量测试方法。对于感知/性能型的非比较性宣称，采用感官描述小组的评价报告或消费者对产品引发感知的自我评价报告支持宣称。此类数据同样要经过统计分析。

6.4　比较性感官宣称

比较性宣称是指将 2 个或多个产品相互比较后得出的宣称。这种比较可以是广告商产品与一个或多个竞品的比较，也可以是同一产品不同版本间的比较。比较的内容包括：优越性、等效性或非劣效性。比较性宣称能突出不同产品在感官特性、情感响应以及产品性能上的差异，或者主要突出产品的新感官特性。

如果某广告商想对公司内的 2 款产品进行比较宣称，例如产品的改进版和当前版，这种挑战风险通常很低。但是，有些公司仍希望获得一些感官数据支持即使风险很低的宣称，因此需要开展测试，对强调新产品与当前产品存在感官特性差异的宣称进行支撑。这种情况下，可使用前期产品的历史测试信息得到 1 个达成标准。由于挑战风险很低，因此选择达成标准时无需像高风险宣称那么严格。

与竞品公开比较的宣传方式会增大挑战风险，竞争对手能对比较性宣称的内容提出质疑，因此广告商要开展严格的测试来支持所提出的宣称，这一点非常重要。

产品优越性宣称是声称广告商的产品更受欢迎或更被偏爱，具有更多让人满意的产品特性或优于宣称中所指的产品。

示例 1：“相比竞品 X，消费者更偏爱我们的品牌”，“品牌 X 的清洁效果比其他领先品牌更好”。

等效性宣称是声称广告商的产品与宣称中提及产品一样好。

示例 2 “味道和……一样好”、“清洁效果和其他领先品牌一样”。

非劣效性（或卓越性）宣称是声称广告商的产品至少与宣称中提及产品一样好。

示例 3 “没有什么产品比……更受欢迎”，“没有什么产品比……更有水果味”，“你无法打败……的美味”，“没有什么比……清洁效果更好”。

等效性宣称是指 2 个或多个产品是相同的。等效性宣称相关的统计假设通常包括上限和下限（即双边备择假设），而与优越性和非劣效性宣称相关的统计假设只涉及下限（即单边备择假设）。所有比较性宣称都有自己的测试设计和分析方法。需要注意的是，当测试结果不具有显著性时，不能推断出等效性和非劣效性。与旨在显示产品优越性或非劣效性的宣称测试相比，等效性宣称需要更多的样本量。

所有类型的比较性宣称（优越性、等效性或非劣效性）都需要经过统计分析处理的定量测试数据。

开展情感比较性宣称时，收集的数据通常是喜好数据或偏好数据。对于感知/性能特性的比较性宣称，既可使用消费者也可使用经过训练的感官小组进行测试。

7　方法

7.1 很多感官分析方法可用于证实感官宣称。依据宣称的类型和测试目的选择支持感官宣称内容的测试方法，具体参见 ISO 6658。

描述性检验和差别检验等感官分析方法通常是以产品为中心，检验目的是确定产品中存在哪些感官特性以及特性上是否存在感官差异。使用经过训练的或有经验的人员担任描述性检验和差别检验的评价员。对于希望建立产品宣称的感官研究人员，非常重要的一点是了解宣称所在国是否有与感官特性宣称相关的特定法规是非常。如果宣称是面向终端消费者/产品用户时，必须是消费者能感知到的感官特性。

开展消费者喜好或偏好等情感测试时应选择合理的、具有代表性的用户样本，如果是市场新产品，选择潜在用户。情感型宣称是基于人的声称，例如“与竞品相比，消费者更偏爱这种产品”。情感型宣称是参与测试消费者响应的表述，宜选取具有代表性消费者样本进行情感测试，消费者样本量宜足够多且招募要求宜相关。

情感测试的消费者包括：

a) 产品的使用者或购买者，或潜在的使用者/购买者（如果产品尚未上市）；

b) 代表更广大的潜在消费者、使用者或购买者群体的测试人员；

c) 对产品主观响应的提供者，例如对产品的喜好或偏好，或采用所有适合项勾选法(Check all apply analysis, CATA) 对感知到的产品特性进行评级或自我报告。

与以产品为中心的感官分析方法一样，盲测和独立判断也是消费者测试方法的重要特征。外部有效性，即测试结果在多大程度上能被推广到更大的用户或潜在用户群体，以及测试产品在多大程度上如同消费者在日常生活中使用的那样，是消费者测试方法中最重要的一点。

分析型感官检验和消费者测试中的测试产品均宜代表市场上可获得的产品，最直接的方式是从零售店购买产品。如果产品还未在市场上推广和使用，那研究者必须提供宣称证实中所用的测试产品与市面产品是一致的的证据。

如果使用一种新的测试方法（例如用于新产品研发）来支持感官宣称，测试方法宜遵循多数感官专业人员所认可的科学性原则。

7.2 探索性、假设性或发现性的研究方法不宜作为感官宣称的主要支持证据。在没有其他证据支持宣称的情况下，不宜使用定性测试方法支持宣称。

7.3 下列方法均可用于感官宣称证实，包括：

a) 整体差别检验（例如三点检验、二、三点检验）；

b) 特性差别检验（例如定向差别检验、感官特性评级法）；

c) 描述性分析；

d) 消费者情感或性能测试。

7.4 用于感官宣称证实的感官测试方法的指导原则参见第五章，具体内容如下：

a) 明确宣称所在国对宣称的相关要求；

b) 主要宣称内容的预定义；

c) 公平的数据收集和公正的测量方法；

d) 与宣称相关的参与者和产品要求。

7.5 宜选用产品的当前用户或潜在用户参与产品情感宣称测试或感知功能测试。产品特性或性能宣称可采用经过培训的感官评价员或优选评价员进行评价。

附 录 A

（资料性）

感官宣称示例

A.1 非比较性——情感宣称：“味道很棒”

一家沙拉酱生产公司希望在产品标签正面印上“味道很棒”。感官、营销和法规团队同意进行产品测试。如果大多数消费者在品尝产品后都认为产品“味道很棒”，公司会将这句话印在包装标签正面。研究人员查阅了消费者评价“味道很棒”的类似产品历史数据，获取相关背景信息以确定决策标准。最终，研究人员提出了决策标准“70%或更多的消费者在品尝沙拉酱后认同味道很棒的声称”。

共计 120 名消费者采用盲测方式品尝了该产品，并对产品声称“味道很棒”的认同程度进行评价。采用七点李克特标度进行评价，“7”是“完全同意”，“1”是“完全不同意”，“4”是“既不同意也不反对”。研究人员推荐使用最高的 2 个级别，即以选择了“完全同意”和“非常同意”的消费者数量作为认同沙拉酱“味道很棒”的消费者数量。这种决定是为了确保只有非常认同的消费者才符合宣称决策标准中的消费者要求。

测试结果以列表形式呈现，120 名消费者中有 96 人（80%）选择了“完全同意”或“非常同意”这种沙拉酱“味道很棒”的声称。基于二项式分布，95%置信区间下 80%的分布范围为（0.717, 0.867），即认同沙拉酱“味道很棒”的消费者比例为 72%~87%，95%置信区间的下限高于 70%的决策标准。因此，测试结果支持在包装上印上“味道很棒”表述。

受控区域内消费者喜好测试方法，参见 ISO 11136。

注 可采用程序完成计算，如使用 R 中代码行 binom.test (96, 120)。使用 Microsoft Excel 2010 完成计算。[1)]

A.2 非比较性——性能宣称：“无残留”

产品研发人员发现了一种成分添加到护肤霜中不会产生残留感。该公司希望将这种成分添加到他们目前最畅销的护肤霜中，并宣传“无残留”。现在需要感官数据支持新配方护肤霜的“无残留”宣称。感官研究人员建议使用经过培训且考核通过的描述性分析小组对添加和未添加新成分的护肤霜进行评价。测试目标是客观确定这种新添加成分是否具有“无残留”效果。描述小组经过培训后对护肤霜产品开展定期评价。研究团队制定的决策标准是认为护肤霜“无残留”的人数更多，即“无残留”回答数比例应显著高于 0.5。（此处对照数据，不是指未添加新成分产品的数据，而是指被添加了新成分、但被判定“有残留”产品的数据。）

描述性评价小组由 12 名经过训练的评价员组成。评价员使用标准涂抹程序将护肤霜涂抹在前臂上，同时对涂抹护肤霜后的前臂表面进行评价。评价员使用标准用量和标准涂抹方法涂抹 2 种护肤霜，每个前臂各涂 1 种。护肤霜采用三位数字编码进行识别，其中一种护肤霜含有“无残留”成分，另一种则不含。两种护肤霜在香气、外观、质地和感觉上均没有差异。

---

[1] Microsoft Excel 是可商购的产品示例。 提供此信息是为了方便本文件的使用者，并不表示 ISO 对该产品的认可。

3min 后，评价员对前臂的“可见残留物”展开评价。

接下来的一周内再次重复相同程序，以确保“无残留”宣称得到强有力的支持。2 项试验数据见表 A.1。

表 A.1　2 次试验数据结果

判　定	测试 1	测试 2	测试 1+测试 2 的总和
判定为“无残留”的总人数	10	8	18
判定为“无残留”的比例	0.83	0.67	0.75

合并 2 项试验数据后进行分析。添加了新成分的护肤霜，在 24 个受试者中总共获得 18 个“无残留”的判定。根据二项式定律，获得“无残留”答案的概率为 0.5，24 次中偶然获得 18 次或更多次“无残留”结果的概率小于 0.023。因此，该公司决定在新配方护肤霜包装上添加“无残留”字样，以向消费者强调其优势。

A.3 非比较性——性能宣称：“易烹饪”

一家公司研发了一种“易烹饪”的芝士意面，与该公司现有市面产品相比，新芝士意面所需的烹饪步骤更少。公司跨职能团队就达成标准形成一致意见，即必须让当前品牌的大多数消费者认为新配方芝士意面“易烹饪”。“大多数消费者”被定义为测试中有不少于 80%的消费者认为新配方“易烹饪”。

为了测试这种非比较性的性能宣称，该公司挑选了 90 位经常购买芝士意面的消费者，向每位消费者提供 1 盒无品牌标识的新产品，让他们在家中自行烹饪。消费者按照包装盒上给出的步骤进行烹饪。要求消费者评价这种芝士意面是否“易烹饪”，并独立完成 1 份关于产品烹饪难易程度的电子问卷。采用 5 点标度评价烹饪的难易程度，“5”表示“非常容易烹饪”，“1”表示“一点也不容易烹饪”，“3”表示“既不容易也不难烹饪”。

90 名消费者中有 80 名（89%）消费者认为新配方产品“非常容易烹饪”或“容易烹饪”。根据二项式定律，95%置信区间下 89%认可比例的分布范围为（ 0.81, 0.95）,即 81%~95%，显著高于达成标准中要求的 80%（p=0.035）。该结果满足达成标准。市场部可以开始准备带有“易烹饪”宣称的包装和销售材料，为后期营销做准备。

A.4 比较性——情感宣称：“减盐 30%，同样美味”

一家薄脆饼干公司希望公司饼干产品更健康，以满足消费者追求更健康零食的愿望。公司决定将其核心薄脆饼干产品中的盐分减少 30%，同时尽可能保持饼干的原有味道。与当前使用的食盐相比，减盐饼干中使用了一种新食盐产品，这种盐颗粒表面积发生了改变，可以提供更强的咸感。现在需要感官数据确定减盐饼干和当前饼干的咸度是否相同。为了满足这个需求，感官研究人员计划采用定向比较检验。

公司跨职能团队同时希望获得消费者数据来支持“同样美味”的宣称。团队认可的达成标准是至少有 70%的消费者在品尝减盐饼干后认同“美味”的声称，从而支持“盐分减少了，但同样美味”的宣称。经过几轮配方改进和内部测试小组测试后，产品研发团队研发了 1 款新配方。与现有饼干相比，新配方饼干具有相近的咸味感知。新配方的盐含量降低了 30%，但感知到的咸度变化很小，同时具有与当前产品类似的滋味剖面。

为了确定减盐饼干咸度是否不低于当前饼干产品，招募了 68 名经验丰富的评价员进行相似性定向成对比较。目的是证明尽可能“少的”评价员（如果有的话）认为减盐饼干的咸度低于当前饼干。团队制定的达成标准是不超过 20%（p_d=0.2）的评价员能够识别出咸度差异，例如正确回答数的比例低于 p= 0.6。

68 名评价员中有 30 名（46%）选择了当前饼干比 30%减盐饼干“更咸”。这个比例显著低于可接受

的比例 0.6（单边 p=0.006，单边置信区间上限是 0.55）。

团队继续测试，确定是否可以使用“同样美味”的声称。

90 名当前饼干产品的消费者采用盲评方式品尝了低盐饼干和当前饼干，并使用 5 点标度分别评价 2 种饼干是否“美味”。选择“4”和“5”才被认为是认同“美味”，达成标准是至少有 2/3 的消费者认同“美味”声称。测试采用单个产品评价方法，即仅使用低盐饼干数据，不使用当前饼干的数据。

90 名消费者中有 65 名（72%）给低盐饼干打分不低于“4”。二项式单边检验表明该比率并没有显著高于 2/3=0.667（p = 0.157，置信下限为 0.63）。该团队得出结论：“减盐 30%，同样美味”的宣称不受支持。因此，新配方低盐饼干不支持“减盐 30%，同样美味”的宣称。

A.5 比较性——卓越性——性能宣称：“清洁效果与领先品牌相媲美”

一家洗涤剂生产商旗下的 A 品牌洗涤剂产品在全国市场上进行销售。A 品牌产品的竞争威胁是 B 公司一个更昂贵洗涤剂，其宣传具有“B 品牌的清洁效果优于其他所有领先品牌”的优越性能。品牌 A 和品牌 B 的销量占全国洗涤剂销售的 91%。

A 品牌公司的研究人员认为 A 品牌洗涤剂和竞品 B 具有相同的清洁效果，该公司希望将该宣称印在产品宣传材料上展示给零售商，从而保持 A 品牌的零售市场份额。同时，公司还希望在电视广告上宣传“A 品牌的清洁效果与领先品牌相媲美”，以对抗 B 品牌的优越性宣称。

A 品牌的达成标准如下。

a) 在实验室中使用标准污垢测试，专家评价员使用 10 点清洁度标尺对清洁效果进行评价，得出 2 种洗涤剂的清洁性能是相似的评分差异不能超过 0.5 分。

b) 消费者在家中连续 2 周使用 A 品牌和 B 品牌洗涤剂后，对洗涤剂清洁效果进行比较，使用 10 点清洁度标尺时，评分差异不能超过 0.5 分。

在实验室准备相同污垢的标准衣物样品，对 A 品牌与 B 品牌的清洁性能进行评价。先使用标准使用量的 A 品牌洗涤剂清洁一批衣物上的污垢，再使用相同标准量的 B 品牌洗涤剂处理具有相同污垢的另一批衣物。公司的 10 名专家评价员在不知道使用了哪种洗涤剂的前提下检查清洗后的衣服。B 品洗涤剂牌的清洁度均值为 8.1 分，A 品牌的清洁度均值为 7.9 分，成对评分的标准误差为 0.4。等效性检验中，两个单边检验 (Two one-sided tests, TOST)和等效裕度δ=0.5 时，p 值为 0.0201（<0.001）。基于 TOST 方法的 2 个 p 值均低于 5%的显著性水平，感官团队得出结论是 2 种洗涤剂的性能非常相似。

150 名消费者采用居家方式使用 A 品牌和 B 品牌的洗涤剂。选择 2 个品牌洗涤剂的实际用户作为测试的代表性用户。用户首先连续 2 周使用其惯常使用的 1 种洗涤剂，并完成洗涤剂性能评价的简短问卷。在使用了惯用洗涤剂一周后，消费者会收到第 2 种洗涤剂，继续居家使用 2 周，并完成第 2 份洗涤剂性能评价问卷。产品使用顺序遵循平衡原则，并采用三位数随机编码。消费者对使用 A 品牌洗涤剂清洗的衣物整体清洁度评分均值为 7.4（满分为 10 分），B 品牌的评分均值为 7.7，差异的标准误差为 1.4。如前所述，使用成对 t 检验的 TOST 方法的等效裕度δ= 0.5，得到 p=0.041 和 p<0.001。由于 2 个 p 值都低于 5%的显著性水平，基于 TOST 方法（不需要多重校正），团队得出结论是基于预设的非劣效性标准，A 品牌具有与 B 品牌相同的性能。

A 公司的研究人员最后得出结论 B 品牌并没有表现更优。同时，A 公司决定通过销售文案、路演和电视媒体等多种形式向零售商传达“没有其他品牌衣物洗涤剂的清洁效果能更优”的信息。

A.6 比较性——优越性——情感宣称：“消费者搭配汉堡时更偏爱的番茄酱味道”

一家调味品公司开发并推出了 1 款口味更丰富、比市场上其他品牌番茄酱更百搭的番茄酱。几家大型快餐连锁店都选择这种番茄酱作为汉堡的标配调味酱。调味品公司在这款番茄酱研发期间进行大量消

费者感官测试，认为这款番茄酱的味道比市场上的另一家领先品牌番茄酱的味道更胜一筹。这 2 个品牌番茄酱占全国销售的 87%。

调味品公司决定开展这款番茄酱和另一家领先品牌番茄酱的消费者偏好测试。团队制定的达成标准是，使用具有地域代表性的消费者样本对 2 个品牌番茄酱进行成对偏好盲测，确定这款番茄酱的偏好。使用了 213 名消费者在全国 4 个不同地点使用汉堡包和番茄酱进行测试，使用α=0.05 和 90%（β=0.1）的检验效力来检测单边偏好测试中 p_d=0.2（p = 0.6）的差异。4 个测试地均是主要人口分布的代表性地区，参与测试的都是经常吃搭配番茄酱汉堡的消费者。给每位消费者 2 份样品，每份半个汉堡，涂上不同品牌的番茄酱。2 种番茄酱的品尝顺序遵循平衡原则，询问消费者更偏爱哪种番茄酱的味道。

测试结果表明，共有 123 名更偏爱这款番茄酱，89 人更偏好领先品牌。单边二项式检验 p 值为 0.010，置信下限为 0.52。结果符合达成标准，可以得出消费者更偏爱新研发番茄酱的结论。这家调味品公司决定在印刷品和电视广告中宣称“消费者搭配汉堡时更偏爱的番茄酱味道”。

A.7 比较性——优越性——性能宣称：“比其他同等价位的品牌结块更少”

一家化妆品公司希望推出一款新配方睫毛膏，宣称“比其他同等价位的品牌结块更少”。这款睫毛膏含有一种特殊成分，这种成分能粘附在睫毛上，但几乎不结块。因此，公司希望这个宣称能获得强有力的技术支持。品类销售数据显示，该睫毛膏与 3 个其他领先品牌的睫毛膏价格相当。

该公司制定的达成标准是，一个由 12 名经过训练的评价员组成的评价小组对该公司睫毛膏产品的结块强度进行评价，结块强度应显著低于其他 3 款价格相当的睫毛膏产品。

该公司最终使用了 16 名有经验的消费者为被试者，每位被试者每天按照标准程序涂抹 1 种睫毛膏。评价员对每位测试者的睫毛进行目视评价，并使用 10 点标度对结块强度进行评分，1 为“无结块”，10 为“极其结块”。被试者和评价员都不知道使用和评价的睫毛膏品牌，不同品牌的涂抹顺序遵循平衡原则。测试期为 3 天，评价员均在涂抹睫毛膏的 2 小时内对每位被试者进行评价。结果如表 A.2 所示。

表 A.2　评价结果

参　数	新睫毛膏	品牌 B	品牌 C	品牌 D
结块强度均值	1.8	5.4	3.8	6.2
p 值（与新产品的比例比较）	–	0.004	0.022	< 0.001

对结块强度数据进行三因素方差分析，包括产品、受试者和评价员（$F_{产品}$= 3.4，df=3 和 738，p=0.017）。进一步，采用单边 t 检验及事后检验对新产品的结块强度均值与其他 3 种产品进行两两比较，结果表明该公司新睫毛膏的结块强度显著低于其他 3 种品牌产品。

根据研究结果，该公司决定针对 3 个竞品提出宣称。

注：此处不需要对多重性进行修正，因为只有所有的成对比较结果都具有统计显著性时，才会进行类别范围的宣称。这与公司针对三个竞品进行测试，并根据测试结果决定具体针对哪些竞品提出宣称的情况不同。该情况下是要根据显著性结果作出宣称，因此需要多重校正，例如 Dunett’s 的多重比较。

A.8 非比较——特性宣称：“苦味更少”或“烘焙味更浓郁”

一家咖啡公司改变了加工工艺，生产出一种苦味和酸味更少、烘焙风味更浓郁的咖啡。该公司希望有证据支持其在包装上标识“苦味更少”或“烘焙味更浓郁”的声称，从而向消费者强调这些变化。感官研究人员向团队建议，描述性分析小组精通评价不同咖啡的感官特性。支持该宣称的达成标准是，训练过的描述小组能证实采用当前工艺和新工艺生产的咖啡在苦味或烘焙味强度上确实存在差异（符合宣称内容）。同时使用 15 cm 线性标度测定时，强度差异均值至少在 1 cm 以上，早期研究中认为 1cm 是保证

消费者具有明显响应差异的阈值。整体显著性水平设为 5%。

感官研究中采用了 11 位经过训练的评价员评价 2 种咖啡（当前工艺和新工艺）的 5 个关键特性：咖啡的整体印象、烘焙味、酸味、甜味和苦味，支持宣称的测试重点是烘焙味和苦味。咖啡喜好和感官评价的早期研究表明，这 5 个关键特性与消费者对咖啡的喜好最为相关。2 种咖啡以标准方式制作，评价员使用 15cm 线性标度评价每种咖啡 5 种特性的强度，从“无”到“极其”，采用盲评方式，重复 2 次。表 A.3 中列出了强度均值及其单边 *p* 值。

表 A.3　强度评分均值及其单边 *p* 值

参数	咖啡整体印象	烘焙味	酸味	甜味	苦味
新工艺制咖啡	8.8	7.9	4.7	5.6	7.1
当前工艺制咖啡	8.1	6.4	6.1	5.3	7.9
p 值	0.092	0.016	0.006	0.538	0.037
p 值（Bonferroni 校正）	_	0.032	_	_	0.074

将产品、评价员以及产品和评价者的交互作用作为自变量，对数据进行方差分析，在 5%显著性水平下对产品进行单边检验。由于宣称是由检验结果决定的，因此需要对多重性进行校正。采用 Bonferroni 校正对烘焙味和苦味这两个关键特性进行校正。表 A.3 中提供了校正后的 *p* 值（乘以 2，因为测试宣称中的 2 个特性）。或者，将主要关注的 2 个（未校正）*p* 值与减半的显著性水平 0.025 进行比较。

结果表明，新工艺咖啡的烘焙味强度显著更高（强度差异均值为 1.5 cm）。但是，多重校正后新工艺咖啡的苦味并不显著低于当前工艺咖啡（校正后的 *p* 值=0.074>0.05，未校正的 *p* 值=0.037>0.025），强度差异均值仅为 0.8 cm。因此，不能支持“苦味更少”的宣称，但可以支持“烘焙味更浓郁”的表述。营销部门开始准备带有“烘焙味更浓郁”的新包装。

附 录 B

（资料性）

I 类错误率与检验统计量的函数

表B.1给出了α（I类错误）是如何随同研究中产品或特性的数量增加而增大。随着检验统计次数增加，错误率随之增大，即观察到至少一个假阳性（显著）结果的概率会随检验次数增加而显著增大。该情况下，当现实中某一特性并不存在差异时，更有可能得出存在差异的结论。

表B.1给出了“多重性”定义的具体说明[7]。支持宣称的描述性分析人员宜意识到这种影响，例如，5%显著性水平下检验10个产品或特性，其中，统计显著性水平下偶然至少1次的概率（即使产品检测特性都是相同的）约为40%。

化解多重效应的最佳方法是在实施研究之前确定产品的可重复性或可靠性或特性差异并严格确定主要目标，或提供其他证据来支持产品或特性的差异。需注意的是，无论何时执行多重检验，均会出现相同效应，包括多个成对产品比较、多个子群细分（例如年龄、性别）或相同产品的多个特性比较。

表 B.1 是使用R3.6.3中的下列代码行所得到的结果：

cbind (x<-1:30, round (1-0.95^x,2), round(1-0.99^x,2))。

表 B.1 概率随检验（产品或特性）数量增加的变化

检验统计量（针对产品或特性）	每组比较的显著性水平		检验统计量（针对产品或特性）	每组比较的显著性水平	
1	0.05	0.01	16	0.56	0.15
2	0.10	0.02	17	0.58	0.16
3	0.14	0.03	18	0.60	0.17
4	0.19	0.04	19	0.62	0.17
5	0.23	0.05	20	0.64	0.18
6	0.26	0.06	21	0.66	0.19
7	0.30	0.07	22	0.68	0.20
8	0.34	0.08	23	0.69	0.21
9	0.37	0.09	24	0.71	0.21
10	0.40	0.10	25	0.72	0.22
11	0.43	0.10	26	0.74	0.23
12	0.46	0.11	27	0.75	0.24
13	0.49	0.12	28	0.76	0.25
14	0.51	0.13	29	0.77	0.25
15	0.54	0.14	30	0.79	0.26

参 考 文 献

[1] ISO/IEC Guide 51:2014, Safety aspects - Guidelines for their inclusion in standards

[2] ISO 6658, Sensory analysis - Methodology - General guidance

[3] ISO 11136, Sensory analysis - Methodology - General guidance for conducting hedonic tests with consumers in a controlled area

[4] ASTM E1958, Standard Guide for Sensory Claim Substantiation

[5] Declaration of Helsinki. Bulletin of the World Health Organization. 2001,79(4), p. 373

[6] CoRBIN R.et al.A Practical Guide to Comparative Advertising: Dare to Compare. Elsevier, 2018

[7] MARTIN G. Munchausen's statistical grid, which makes all trials significant. The Lancet. 1984, 22;2(8417-8418),p.1457

[8] Advertising Standards of Canada. Guidelines for the Use of Comparative Advertising: Guidelines for the Use of Research and Survey Data to Support Comparative Advertising Claims, 2010

[9] SCHNEIDER-HÄDER B., HAMACHER E., BEEREN C. Sensory Claims - Methodological approach to development and substantiation. DLG-Expert report 15/2015. DLG e.V, Frankfurt, 2015. Available from: https://www.dlg.org/fileadmin/downloads/lebensmittel/themen/publikationen/expertenwissen/lebensmittelsensorik/e_2015_ 15 Expertenwissen SensoryClaims.pdf

[10] ASAI. Code of Standards for Advertising and Marketing Communications in Ireland. Advertising Standards Authority for Ireland (ASAI), Dublin, Ireland

[11] NAD:http://www.asrcreviews.org/asrc-procedures

[12] UK. Advertising Standards Authority (ASA) and Committees of Advertising Practice (CAP) https://www.asa.org.uk/news/a-quick-guide-to-comparative-advertising.htmlt guidelines:_yVDGXyCy7-I with specific codes at: https://www.asa.org.uk/codes-and-rulings/advertising -codes.html

[13] NBC Universal Advertising Standards. Advertising Guidelines: https://nbcuadstandards.com/files/NBC Network Advertising Guidelines.pdf

[14] ABC Television Network. Advertising Standards and Guidelines: https://abcallaccess.com/app/uploads/2016/01/2014-Advertising-Guidelines-.pdf Maueswe. ass

[15] Committee of Advertising Practice (CAP): www.cap.rg.uk