监管决策中的信息质量——同行评审与良好实验室规范之比较

Lynn S. McCarty,[1] Christopher J. Borgert,[2,3] and Ellen M. Mihaich[4]
[1]L.S. McCarty Scientific Research & Consulting, Newmarket, Ontario, Canada; [2]Applied Pharmacology and Toxicology, Inc., Gainesville, Florida, USA; [3]Center for Environmental and Human Toxicology, Department of Physiological Sciences, University of Florida College of Veterinary Medicine, Gainesville, Florida, USA; [4]Environmental and Regulatory Resources, Durham, North Carolina, USA

背景：有关如何最好地确保毒性实验数据来源的有效性和可信度的讨论正在进行中。其争论的中心是期刊同行评审程序是否优于为符合法规要求而采取的良好实验室规范（GLP）标准。

目的：我们试图评估基于同行评审程序与GLP标准的监管决策的理论基础。

方法：我们研究了已发表的关于同行评审、GLP规范以及法规制定中如何评估科学数据质量与有效性的相关文献。

讨论：有观点认为同行评审是为实验数据生成、分析和报告提供质量控制的连贯、一致的评价程序，能够充分可信地确立其相对优势，而GLP仅被看作旨在防止研究者舞弊的跟踪程序。已发表的分析文章显示同行评审过程有主观性与多变性，因而不支持以上观点。尽管GLP并非旨在确立相对优势，但它是一种国际公认的实验行为与数据记录的质量保证与质量控制方法。

结论：这两项程序都不足以完全确立相对的科学合理性。然而，在同行评审程序与监管指导中科学信息交流更清晰、更透明的变化，显示出在确保信息质量上新出现的趋同倾向。确定何者更具优势的解决方案，在于建立一种证据确凿、普遍认可的证据权重方案，在监管决策中对同行评审与GLP信息均进行评估，结合两者的优势和特定的相关性来指导决策过程。

关键词：数据质量；GLP良好实验室规范；同行评审；监管决策；毒性试验。

Environ Health Perspect 120:927–934 (2012). http://dx.doi.org/10.1289/ehp.1104277 [Online 17 February 2012]

科学数据的有效性和可信性对所有科学工作以及利用这些数据而做出的决策框架都极为重要（Schreider等 2010）。其中最重要的是由美国的联邦机构，如美国环境保护署（EPA）、美国食品和药物管理局（FDA）、美国农业部(USDA)或其他行政管辖区的类似机构例行做出的风险评估、安全性评估及监管决策。但由于其依赖的数据类型或来源均接受过严格审查，这些监管决策常受到质疑。监管决策受挑战是因为其依托的数据，据称对保护公众健康和环境缺乏相关性或敏感性，这些数据是由被认为与决策结果有利益冲突的科学家或实验室得到的（例如Myers等 2009）。有些提议的解决方案支持透明化，并强调原始数据与方法学细节的可用性作为提高公信力的主要方法（Borgert 2007；Schreider等 2010）。

透明度越高越会提高决策的公信力，因为它增强了程序的诚信感。另一方面，透明与诚信本身并不解决数据质量的根本问题。科学期刊的同行评审以及对法规制度数据的接受都认可严格评估数据质量的必要性，然而解决这一问题的做法和程序却因处理科学数据机构的不同而存在差异。这种差异可能因数据质量的定义不同而引起，但更可能与裁定数据质量的理由有关，而理由因机构的权限范围而异。在本综述中，我们对假定为监管机构认可与发表在同行评审期刊上的数据质量的不同定义进行比较和对照。然后，我们分析这些不同标准如何应用于环境卫生与安全性监管决策中，重点放在每种标准如何提高或损害决策过程的可信度。为了说明有关问题，我们参考目前内分泌干扰物研究中关于这一问题的争论。

自20世纪90年代“内分泌干扰物”或“激素活性剂”（National Research Council 1999）概念形成以来，就一直对此争论不休。环境中低水平的各种有机化学物可能对人类与动物造成不易察觉但广泛存在的发育与生殖影响的理论，源自Wingspread会议和名为《失窃的未来》(*Our Stolen Future*)一书（Colborn等 1996；同见vom Saal 1995）。为更好地鉴定低水平暴露可能的影响性质与程度，科学界的反应是对化学物、生物及反应终点进行复查。这包括延伸与扩充现有的生殖、行为和生化方法及终点，以及检测新方法与新终点。各种国内与国际监管及咨询机构已采取全面测定方法评估潜在问题的性质和程度，并确定如何将评估结果纳入目前人类与环境法规的最佳方式。技术与政策的挑战是强大的。尽管经过相当大的努力，标准化技术定义的争议依旧，普遍认可的测试方案的广泛应用被证明是个尚未解决的重大挑战（Borgert等 2011b）。

在内分泌问题上的争论，大部分涉及毒理学测试来源的重要性以及研究者可能的偏见。由于对企业资助的明显偏见，有些人认为受企业委托的毒理学研究质量和可靠性较差（例如Sass等 2005），受企业雇佣或其研究得到企

业资助的科学家受到腐蚀，不符合政府政策或技术小组的工作要求[详情与反驳见Barrow和Conrad（2006）]。由于政府授权的许多研究既接收大量的企业资助也有政府工作人员参与，照此逻辑，意味着按照这一政策，应取消许多政府科学家的资格。同时，这一观点忽略了这样一个事实，即非企业科学家也有与资助有关、影响其观点的偏见，因而按同样逻辑也应取消资格。科学界的许多人士反应强烈，提出“判断科学，而非科学家”的观点（例如Borgert 2007；Gori 2009；毒理学学会 2008）。

数据出处与偏见成为一种广泛应用的化学物——双酚A的争论中心。美国和欧洲的监管机构主要根据良好实验室规范（GLP）进行的几项啮齿类动物研究的毒理学结果，认为目前人类的双酚A暴露有足够的安全边际［即欧洲食品安全局（EFSA）2006；FDA 2008］，但研究和监管活动仍在继续（FDA 2012）。Myers等（2009）认为在监管决策中不应优先采用根据美国EPA或经济合作与发展组织(OECD) GLP指南进行的研究，因为在他们看来，GLP程序不如用于决定科学期刊接受发表的同行评审过程严格。接踵而来的是附和（Becker等 2009；Tyl 2009）、补充说明（vom Saal与Myers 2010）及反驳意见（Becker等 2010；Tyl 2010）。

尽管期刊同行评审与GLP成为以上争论的中心，却一直未从各自在不同学科中的适当性，对监管/法律用途的影响上，就它们的过程和目标间相似性和差异方面进行过深入分析。作为第一步，我们从实验数据、数据分析以及科学出版物和政府报告注释的质量保证和质量控制方面，调查期刊同行评审程序与GLP的现况。然后，我们将讨论扩大到对同行评审、GLP中科学数据的质量与有效性及监管发展的评估。

阐明同行评审的定义是必不可少的。同行评审常指科学期刊在筛选发表文章时所使用的程序。但是，在其它领域，包括对研究合同/奖学金和私人及政府科学报告、科学政策文件以及监管指令的评估中，也使用类似的方法。美国EPA所使用的科学顾问委员会或小组（SAB / SAP）评审与评估是个很好的例子。这种非期刊同行评审与期刊同行评审共同面临的问题是：由于缺乏一个证据充分、广泛适用以及普遍接受的同行评审方案，评估过程与行动及其后随访的性质、程度和彻底性差异很大。因此，尽管许多有关同行评审的公开辩论重点集中在期刊同行评审上，但我们的研究属于较广义的同行评审活动。

相关文献的评价

同行评审的历史与发展：Burnham (1990)、Kronic (1990)和Rennie (2003)主要根据生物医学的经验，发表了有关科学同行评审历史的优秀综述。从18世纪初同行评审正式开始到20世纪40年代中后期的现代化发展阶段，有两个重要特点。首先，主要根据责任编辑的偏好，其性质和范围在期刊内与期刊间的差别很大。其次，同行评审根据对期刊读者的相关性和潜在意义，帮助选择发表的内容与议题。同行评审与发表并不保证提出的理论、数据、分析或结论的有效性或正确性。

二次世界大战后，由于科学研究的剧烈增加，期刊数目扩大以适应科研的发展。到20世纪80年代情况开始发生改变。尽管期刊同行评审的成长和发展壮大，整体格式却一直不统一，评价的范围和性质仍千差万别。历史悠久的《英国医学杂志》(*British Medical Journal*)的编辑Stephen Locke出版了一本有关同行评审的重要著作（Locke 1985）。至少部分原因是对Bailar和Patterson（1985）一篇文章的回应，《美国医学协会杂志》(*Journal of the American Medical Association*)于1986年举办了4年一次的关于同行评审研究系列会议的首届会议：同行评审与生物医学出版物国际会议。尽管发表的同行评审评价/述评有增无减，但同行评审研究作为公认的科研课题此时才刚开始。Campanario（1998）曾对这期间前二十年的同行评审研究的范围和多样性进行过调查。

关于同行评审研究文献的综合评价尽管非本综述内容，但我们在此简要概述一些重要的论题及争论。最近有些舆论/述评谈到同行评审的普遍性问题并表达了对其性质与范围的关注：“同行评审垮了吗？”（McCook 2006），“同行评审寿终正寝了吗？”（Henderson 2010），“我厌恶你的论文”（Akst 2010）。不过，这些关注是长期存在的。Horrobin（1990）认为，同行评审的被拒必定继续对挑战知名科学家控制大部分评审现状的新发展起阻碍与削弱的作用。Armstrong（1997）同意这一观点。Jefferson等（2002）备受关注的综述的结论认为，同行评审对科学出版物质量的影响基本上未经验证，主要是因为它没有明确而一贯的宗旨。

2006年《自然》杂志发起了网上同行评审讨论，引发了一些评论文章，其中有以下几篇。Jefferson（2006）指出，同行评审可能是可利用的最好程序，但只有经过对各种选择进行系统评估后才能确证这点。Jennings（2006）认为同行评审效果的量化指标越多，越能更好地评估现行做法和各种选择。Lee和Bero (2006)主张改进透明度和公平性政策并简化程序。

尽管对期刊同行评审的目的、目标和流程普遍存在质疑，却并无减少甚至取消同行评审的广泛运动，相反却出现了许多如何改善和加强期刊同行评审的建议。Armstrong（1997）讨论了从本质上改革同行评审的各种方式，从是否发表的二分法决策到以何种形式发表文章的分级方案。Benda与Engles（2011）指出，当将同行评审视为一种判断性预测任务时，它是比较成功的，但可能会妨碍创新型研究工作的发表。他们提出补救措施以增加创新性研究的发表机会，其中

涉及改革评审员的表决与决策程序。

Cicchetti（1997）呼吁为同行评审决策的规范化制定明确的决策准则。另一些人则将重点放在数据报告的具体方面。Kilkenny等（2010）主张采用“动物研究：体内实验报告（Animals in Research: Reporting *In Vivo* Experiments，ARRIVE）”的指南来报告保健科学领域的动物实验数据，指出试验报告统一标准(Consolidated Standards of Reporting Trials, CONSORT)指南，在提高随机临床试验报告的数据质量和透明度上所取得的成功。Borgert（2007）提出对同行评审出版物的要求，包括综述文章遵循临床期刊系统评价所采用的文献检索与选文规则，以及所有研究论文在线发表包括原始数据、实验室记录本与统计算法在内的详细细节。然而，即使这种类似GLP的建议也并未解决众所周知的期刊偏见，即倾向于发表明有显毒效应的研究。对于新方法或老法新用的研究尤为如此，通过展示其实用性来赢得关注和接受。总之，以上因素使效果很小或无效，即“负结果”的研究被低估，因而使许多毒理学问题难以得到公正的看法。

同行评审和生物医学信息在法律/诉讼中的应用是一个重要的分课题。美国最高法院裁定，尽管同行评审可提供关于科学价值的有用信息，但它并非衡量有效性的绝对可靠的标准，也并不总是值得信赖（Horrobin 2001）。按照Horrobin的观点，这点以及其他最新研究的结果要求要么改进，要么抛弃同行评审。Henry与Conrad（2008）回顾比较了以监管倡议或以一般科学为目的所进行科研工作的评判问题，结论是美国的司法和行政指导批准了一项通用评价方案，该方案与信息生成的出处、资金来源或研究背景无关。Boden与Ozonoff (2008)研究了这一问题，结论是各种利益冲突并非专门针对为诉讼产生的科学问题，相反，冲突本身在本质上是普遍存在的，在科学中也无处不在。他们还指出，不应将同行评审当成判断质量和相关性的一个可靠指标。

法律科学工作的潜在利益冲突问题达到了炽热化，为此毒理学会专门在其“1998年毒理学研究优先领域的原则”（1998 Principles for Research Priorities in Toxicology）中(毒理学学会 2008)纳入了：

4. 应该根据科学价值评判研究工作，而不考虑其资金来源或研究在何处进行(例如学术界、政府还是企业)。

总之，虽然许多人都认为科学方法与数据的有效性在决定发表与资助价值上居于首位，不应受背景或经济利益的限制，但对其公开的要求却有着各种观点。有些人承认，经济利益的公开可能会限制某些方面数据解释与利用的分享（Barrow与Conrad 2006；DeAngelis与Fontarosa 2010），另一些人（Borgert 2007；Gori 2009）则认为，由于绝对没有利益冲突是不可能的，除了公开财务状况和背景以外，应要求数据和方法完全透明。

发表文献中的欺诈行为已成为综合研究的最新和引为鉴戒的主题之一。Steen（2011）研究了2000年至2010年间被同行评审生物医学期刊撤回的近800篇论文。尽管疏忽性错误是最常见撤回原因（70%以上），他还发现了作者蓄意欺诈但最终被捉，强行撤回的故意欺骗证据，指出发表在高影响力著名期刊（如《科学》或《自然》等）上是这类欺诈行为一个共同特点。一个突出的内分泌干扰领域的例子就发表在《科学》上（Arnold等 1996）。尽管次年该文被撤回（McLachlan 1997）并最终认定为欺诈行为[卫生与人类服务部（DHHS）2001]，但最初由这篇传说中的“开创性”论文所引发的讨论和关注着实在监管决策和公共政策中露了一手。

即便高影响力期刊在发表前有严格的同行评审程序，但不要期望它能识别所有的故意欺诈和欺骗。美国管理和预算办公室（OMB）对此认同，尽管期刊同行评审有价值，但仍有许多发表有缺陷或伪造数据的例子记录在案（OMB 2002）。发表后又重新发表的尝试通常是利用了不可靠的（评审）方法或蓄意欺骗，由此也凸显了期刊同行评审并非在数据质量或科学合理结论上完全可靠的方式。

总之，期刊同行评审程序的确存在许多问题，但提出的解决方案更多是修正，而不是替代。重要的是，同行评审的目的并不是为了确保研究设计或者数据收集或分析方法的质量，也并不保证解释得到数据的明确支持。相反，同行评审的目的是帮助确保所发表文章有思考与讨论的价值。为此，同行评审考虑的是文章涉及的主题是否适时，是否引起目标受众的兴趣并对他们有意义，方法与结果细节的叙述是否足够详细并能够复制，以及结论是否得到所提出数据(并非唯一或明确证明过)的支持。基于这个局限性，我们不期望期刊同行评审能发现欺诈、不当行为，甚至有某种程度的偏见报告和解释。但是，我们可以鼓励同行评审采取更规范的程序，不单以全面透明地报告数据的生成和分析为重点，而且报告审稿过程本身。这样不仅促进对新发现的公平性，而且更容易识别不当行为或偏见，这将有利于对其他用途，如监管和诉讼进行发表后评价。

GLP的历史与发展： GLP的历史发展有据可查（例如，OECD 1998；世界卫生组织 2001）。简要地说，美国FDA针对实验行为与新药注册提交报告中的各种问题和困难，于1976年提出了GLP法规，并于1978年生效成为《最终规定》(Final Rule)。该法规重点通过规定具体的实验行为和报告条款，提高临床前药物安全性数据的质量。随后美国EPA提出GLP法规，《最终规定》于1983年生效。按经济合作与发展组织GLP原则建立的GLP国际论坛按，以《化学品评估数据互相认可》(Mutual Acceptance of Data in the Assessment of Chemicals)为指导（OECD 1981a，1981b）。此后，OECD GLP指导经过修订，扩大为含各种问题与领域的15

个指导文件系列（OECD 2011），这些文件已为经济合作与发展组织的不同行政管辖区正式接受［例如，欧洲联盟（欧盟）2004］。OECD GLP始终鲜明而明确地关注质量保证与质量控制（QA/QC）(OECD 1998)：

良好实验室规范（GLP）是关于非临床健康和环境安全研究的计划、执行、监控、记录、存档与报告的组织程序与条件的质量体系。

GLP还有一个长期目标：相互可接受数据（OECD 1981b）。相互可接受数据确保收集和报告足够的实验数据，以便不同行政管辖区可利用按GLP进行的研究，以符合当地法规要求。这样限制了潜在的技术交换的障碍，降低了提供化学品监管数据的整体成本，并减少监管活动所使用的动物。

为避免将GLP视为已在全球范围实施并统一强制执行的国际标准，Helder (2008)指出，尽管共同目标是证明这些设备所得到的数据是可靠的，可用于所有行政管辖区对化学品的安全性评估，但OECD成员国间在应用这些测试设备实施GLP检查的方法并不相同。同样，Huntsinger（2008）发现，美国EPA、FDA以及OECD间在GLP实施上有一定差异但并不显著，不会影响到数据的质量或完整性，但提高一致性是一个重要的长远目标。

除高水平OECD、美国EPA和FDA文件提供广泛指导外，还开发了具体项目的详细GLP指导。美国鱼类和野生生物服务(U.S. Fish and Wildlife Service)在非人类药物测试中使用基于GLP的协议［水产动物药品认证合作伙伴（Aquatic Animal Drug Approval Partnership, AADAP）2006a，2006b］。GLP规范的统一指导促使欧盟成员国取得数据的一致性（EU 2004）。欧洲化学品管理署（ECHA 2008）探索GLP相关性和可靠性的全面规范，并要求相关性清单，以及对有关数据可用性和方法描述的类似GLP可靠性问题的答案。

简而言之，GLP起源于美国，并很快被国际公认，通过OECD的支持，成为确保以完全标准化方式执行的实验科学协议以及监管目的的化学测试数据报告的质量保证/控制程序，因而得到的信息被许多行政管辖区的各项工作所接受。不断发展中的GLP规范将继续对科研管理、报告及监管应用起到巨大影响。

讨论

同行评审、GLP的历史背景以及相关科学信息质量改进的努力与实践（例如，Batterman等 1999；Burnham 1990；Kronic 1990；Rennie 2003）提供了解决前文认定的两个关键问题，即研究者偏见与数据有效性的平台。Myers等（2009）认为，GLP规范不应是监管活动中应用的科学信息的金标准，而通过期刊同行评审程序的科学报告才应是金标准。他们的中心论点有两点。首先，他们声称科学同行评审是一个为数据生成、分析和报告提供质量控制的连贯、一致的评价程序，从而为信息相对价值和结论强度的确立提供依据。其次，他们坚称GLP不是同行评审程序，其提供的质量保证/控制不够或较低；因此根据GLP获得的信息与结论较差。下面逐一讨论他们的以上论点。

很难从过去25~30年间所发表的庞大的论文与评论中证明科学期刊同行评审是个连贯、一致、可靠的评估程序。基于前面介绍的概述，相反的结论可能更准确。与GLP正规的QA/QC程序，有定期审查和更新的具体书面目标和指导不同，科学期刊的同行评审具有不同的要求与程序特点。各家期刊自行决定评审员的选择程序、评审员指导（如果有的话）以及手稿被接受发表的程序。随着时间推移与编辑的变化，期刊间与期刊内部的方法会有所不同。这种多样性和严谨性的差异阻碍了同行评审期刊发表稿件连贯、一致评估的程序。此外，与Myers等（2009）的断言相反，临床研究的系统分析已发现顶尖期刊的资金来源与数据的报告质量间没有关联性（例如，Kaiser等 2011）。

在早期正式成文的指导期刊同行评审的例子中，《英国医学杂志》（*BMJ* 2004）的同行评审员培训资料尤为值得一提，它包含背景材料、明确的过程目标、良好的评审范例，以及如何为该杂志评审的详细指导。然而，很难确定数据质量与完整性以及解释的科学合理性这些根本问题，因为几乎所有期刊同行评审都是在得不到基本数据，或常没有实验方法详细信息的情况下进行的。当出版商为扩大读者数和降低印刷成本精选文章时，他们也在不经意间限制了评审员和读者恰当评估科学所需的信息。

Myers等（2009）论点的第二部分是，GLP不是一个同行评审程序，而且不如同行评审。这一点仅在GLP不需要在被评审的论文中在其涉及的研究领域上由知识渊博的几个科学家发表相对自由、保密的评论的意义上才正确。与期刊同行评审不同，GLP从关于收集什么信息，以及如何收集与报告信息方面，对实践者进行明确和详细的“先验性”指导。目前的OECD GLP代表着数以百计（如果不是千计的话）的科学技术专家，对应用GLP程序的同行进行集体指导。GLP规范本身定期进行评价、修订、提高，并不断充实。

由于GLP常用于指导监管法规所要求的毒性研究，它常被误解成指导性研究的代名词。批评者认为指导性研究，由此被错误联想到GLP，可能未纳入某特定领域的最新进展。不论是否是最先进的研究，指导性研究的相关性和可靠性是经过一定程序证明，许多指导性研究经受了正式的验证，包括在多个实验室进行同行评审的环测试，随后对数据与分析进行同行评审。发表在科学期刊上的研究论文由于采用的方法常常太新，尚未经过这种测试，反映出两者不同的侧重点。尽管如此，GLP能够并经常应用于新的探索性研究。

实际上，GLP是实验规划的框架和正规QA/QC程序，需要把做了什么以及如何做有详细的文件记载。需由QA/QC人

员确认规范，提供衡量遵循选定设计的可靠性和有效性的方法。此外，GLP项目及设施须经过正规培训人员的合规性审计。这并不意味着GLP保证实验数据的解释、分析和结论是正确的，或总是使用最能提供证据和最尖端的技术。GLP在解决各种科学有效性问题上并不比期刊同行评审强。然而，它确实在最大程度上达到了保证数据的二级有效性。因此，GLP能对原始数据进行彻底复查和重新评估，无论是核查原有解释还是进行新分析。例如，通过对Pinter等的莠去津研究(1990)的再分析，SAP确定高剂量时出现的雄性乳腺肿瘤发生在寿命明显长于对照的大鼠中［联邦杀虫剂、杀真菌剂和灭鼠剂法（FIFRA）SAP 2000]。原来的同行评审未认识到肿瘤是由衰老，而非莠去津引起。SAP和美国环保署后来承认了这点以及研究设计和数据分析中的其他缺陷（2000）。这种再分析之所以有可能进行，是因为报告要求保留QA/QC，存档以及可获得原始数据。

因此，有关期刊同行评审与GLP孰优的争论，比较的是用于不同目的的不同实体。GLP在服务于某些监管目的时非常好，无疑优于期刊同行评审程序。我们的顾虑是尽管监管评审应包含未纳入GLP的期刊同行评审这一重要部分是合理的，但是这并不排除GLP有明显好处。解决同行评审与GLP孰优的争论反而更需要加强这两项程序。

GLP与同行评审的趋同：GLP与同行评审都对科学报告和评价非常有用。一方面，同行评审的最终目的是确保发表的文章值得科学界关心与讨论，提供各发展领域新的、有意义的、易理解的相关材料。另一方面，GLP的总目标是确保全面、一致地详细报告实验研究的各个方面，从而有可能易于进行再分析和重新评估。尽管GLP为数据收集和报告带来严谨性，但并无必要要求所有科学研究都这样做（例如，Miller等 1999）。有些人认为，严格按照GLP规范的额外成本将非常高昂，学术研究尤为如此。然而，透明度的提高将有助于同行评审人员评估文章发表的总体价值，也有利于他们发现疏忽的错误和蓄意欺诈。

GLP与同行评审在保持它们各自主要目标的同时，在某些方面的趋同似乎是合理的。这种趋同主要在同行评审领域十分明显。有些期刊现在要求报告额外数据，许多期刊提供电子版辅助材料供选择。《英国医学杂志》详细的同行评审培训与文件指导引人注目。有些期刊正在进行改革同行评审方法的尝试，包括优化指导，公开或不匿名评审，以及减少发表前筛选结合发表后公开评论等各种改革。

监管机构对确保科学概念与结论的交流畅通以及数据透明度的提高也越来越感兴趣。继2001年美国颁布信息质量法(Quality Act)后，美国行政管理和预算局（Office of Management and Budget, OMB）发布了“确保并最大限度提高联邦机构信息传播质量、客观、实用和完整性指南”（OMB 2002）。最新的同行评审与风险评估指导继续强调数据质量与交流的畅通（OMB 2004，2007）。其他机构，如美国环保署和美国鱼类和野生动物服务以及DHHS的下属机构［例如FDA和国立卫生研究院（NIH）]已制定了在执行和应用方面的具体指导来修正他们的工作（AADAP 2006a，2006b；Birnbaum与Culpepper 1999；DHHS 2006；EPA 2006）。欧盟的化学物质注册、评估、授权和限制（REACH）程序也制定了解决数据相关性、可靠性和充分性的指导（ECHA 2008）。

风险评估评审指导对方法报告、数据可供查询及质量的要求有助于弥合GLP/同行评审之间的鸿沟（OMB 2007）。在加拿大，科学和技术咨询框架(Government of Canada， 2000)在保护人类和环境卫生法规中写明，在实施与评估科学信息及决策过程中，应遵循政府规定的质量为本的原则。然而，Forristal等（2008）指出，应用这些原则进行化学物风险评估的具体运作框架尚未全部完成。

最后，在Myers等的讨论（2009）中显然没有提到这样一个事实，即对制定监管指导有关的各种来源的所有数据进行的评审均由监管机构自己完成。由于GLP报告特别为便于发表前后的评审而设计，其报告的要求比较严格，因而它们更易于数据质量的重新评估。因此，对GLP研究数据的信任常无可厚非地高于同行评审期刊论文的数据，后者因缺乏报告细节而制约了数据的彻底复查。由于期刊同行评审、GLP和监管决策的数据透明度和质量目标相似，它们间的方法趋同性肯定将会继续，这种趋同性无论对这三种评估活动的内部还是相互间都会有许多好处。

有效性：数据质量、总体研究质量、相关性：Borgert等（2011b）曾描述过科学有效性的三个层次。要使确立的科学事实被承认，科学数据必须至少符合三项基本原则，这三项原则是能真实测量自然世界的科学基本话语的依托（Gori 2009）。这也许可以称之为数据的“主要有效性”。首先，科学测量的同一性和真实性必须在一定的精度范围内可验证。其次，测量和观察必须不混杂有已知会破坏其准确度和精密度的外在因素和影响。第三，测量和观察都必须可被独立的第三方复制。这三项原则是无可争辩的，并批准为美国有效管理科学的最基本要求（U.S. Congress 2010；U.S. House of Representatives 2010）。我们相信它们也为所有评判数据提供了足够明确的主要标准。确立数据的可靠性，还需要数据报告的透明性和彻底性(Klimisch等 1997)，它构成了数据的“二级有效性”。研究的整体相关性与发表价值，也许称之为科学数据的“三级效性”。这三级科学有效性（Borgert等 2011a）包含了科学数据评估的必需要素。GLP和同行评审包含了有效性的许多重要方面，如精度与监管相关性，但都未以明确的方式完全解决所有三个层次的问题。

所谓的“证据权重”（weight of evidence，WoE）评估常用于对为达到监管决策所应用的各类研究的结果进行检查、优先排序与整合。为达到明确鉴别总体研究质量与确定管理决策投入的相对价值的预期目标，需要有针对要解决的假设或问题的明确程序（Borgert等2011a）。klimisch等（1997）定义了完全适用于该任务的可靠性、相关性、充分性，他们还定义了研究/数据的四类可靠性：无限制可靠、受限制可靠、不可靠以及未定。Schneider等（2009）通过开发具有统一客观类别标准的评估工具来使用Klimisch的可靠性类别，使其能对毒理学研究和数据进行科学地合理评估和相对价值加权分配。该评估工具可用于给定情况下的所有研究（包括GLP和非GLP研究）。有人还开发了评估生态毒理学研究数据的类似工具（Hobbs等2005）。在这两个例子中，都是开发了一组问题，指导对已发表和未发表的数据进行科学严谨性分级和评估，以帮助提高评审的一致性与透明度。

Schneider等（2009）与Hobbs等(2005)发现在他们的研究所评估的出版物中，同行评审对报告与解释质量的评估存在可变性。Hobbs等（2005）指出造成同行评审不一致的原因很多，包括未发现报告的数据与解释的不一致性。例如，Hobbs等(2005)讲述道，虽然外部评审员都认为他们评估的研究之一测量了温度，该研究只是说测了一个燃烧室的温度，却未提交测得的读数。Schneider等（2009）还发现，评审员对信息的打分，因其对评估方案提出的问题的解读而异。

如果Schneider等（2009）与Hobbs等(2005)代表了同行评审出版物的同行评审的话，那么稿件质量与完整性的差异则指出了期刊同行评审存在的问题。如果评审员不同意一篇已通过同行评审稿件对具体问题的答案的话，在未提供具体问题或指导时，他们的评审怎么可能全面而可靠呢？至少，像Schneider等（2009）和Hobbs等（2005）提倡的清单将有助于以较为类似GLP的方式指导期刊评审。

然而，不论质量得分有多高，任何研究不会对具体决策有特别大的用处。在期刊同行评审或GLP中增加一项评估步骤，既不现实，也不可行。尽管在证据权重中数据相关性对研究相关性极其重要，但数据相关性评估是在一般相关性概念内，而非特定环境条件下进行。

但是，美国环保署详细描述了具体相关性概念（2006）：

> ***DQA[数据质量评估]是建立在一个基本前提上，即数据质量只有在它涉及数据的预期用途时才是有意义的。数据质量不存在于真空中，评审员需要知道在什么情况下要利用数据，以此作为确立判断数据是否可以接受的参照标准。***

尽管在监管应用中有关相关性和充分性的判断有具体条件的问题，但制定类似Schneider等(2009)和Hobbs等(2005)提出的一般分类与标准，更好地帮助确定具体相关性应该是有可能的。这将为监管决策评估毒理学研究/数据的相对价值提供了更透明和更有效的证据权重方案。这种做法不需要任何同行评审与GLP的论点，重点放在监管决策者所面临的关键问题上：确立某个问题的所有毒理学信息的可靠性、充分性和相关性。一项旨在评价用于监管活动的期刊同行评审与GLP信息，证据充分、普遍接受的证据权重方案，对于期刊同行评审与GLP活动中确定相对价值和一般有效性方案的更新是有帮助的。平衡数据的有效性与具体相关性是一项挑战。

证据权重的提出

关于证据权重的背景资料和普遍适用(也适用于内分泌干扰物筛选)的科学总则(EPA 2011)，以及为具体监管目的得到的假设驱动的证据权重方法的例子(Borgert等 2011a)，将不再在此赘述；然而，一些主要信息在需要时将会被提及。这些基本原则可经扩展开发成普遍适用于监管同行评审的证据权重框架。具体内容可灵活处理，以解释不同的管理目标和用途。为此，必须对“权重”和“证据”的可信度和透明度进行明确定义。

如同前节所述，已按数据的主要、二级和三级有效性（Borgert等 2011a）定义了科学证据（Gori 2009）。这些概念公认为比较确切，与透明报告文献检索和遴选程序的建议相结合，可用于所有毒理学研究的评估。由于毒理学数据和分析常应用于原研究者未预见到的情况和环境下，使用者需要考虑到当初的意图及新提出的用途。如果有可能，应对通过新用途测试的假设进行明确定义。新用途将指导文献检索和选择过程，这点应在证据权重分析前就制定好并记录下来。可在原来意图条件下对主要有效性与二级有效性进行评估，因为测量和报告质量的可靠性不会因新用途而改变。三级有效性评估常需在两种条件下都进行，因为一项证明其原来目的的研究设计可能不包含新用途的关键成分。然而，除了极少数例外，一项原来目的就太弱的研究不可能增强另一项研究的证据强度。

另一方面，“权重”意味着不同数据的价值或重要性不同，因此与“证据”相比，更需要根据具体情况确定“权重”。在加权步骤中，使用者必须仔细考虑到预期的监管用途，而不是研究者的原有重点。理想情况下，目标一致时可定量分配权重（Borgert等2011a），对诸如预测力和假阳性/阴性检出率等因素作出解释。反之原来目的与监管目的可能不同，定量基础常难以达到，灵活性对于广泛适用于各种监管活动是必不可少的。灵活性可能允许明确纳入一些规定，以抵消对阴性毒理学数据的发表偏见，这种偏见对最新方法学或现有方法的新用途可能尤其不确定。尽管如此，对成功和普遍可接受的证据权重计划，有两个因素至关重要：*a*）用于各类数据，包括其文献基础加权的过程，必须是透明和密切联系的，和 *b*）权重本身必须通过推理得到，并

始终如一地执行。

本综述无法囊括全部例子，不过用于监管目的的具体数据关联性评估指导(U.S. EPA 2006)提出了鉴别与加权具体应用数据的程序。然而，在我们看来，一个可信的证据权重评估方案除叙述一般原则外，还必须包括具体标准与应遵循的步骤。不论数据是否按物种、暴露途径、分析方案、试剂级别、剂量的药代动力学评估、现场与实验室或任何其它参数优先分级，每项加权的合理性都应通过广征博引的解释予以证明。正是在这一为特定目的信息加权的过程中，解决了GLP与期刊同行评审的争论。对需要数据再分析的监管用途来说，GLP样特点可能至关重要。对于其他目的，使用最敏感和最新分析技术得到的数据可能会占优势。应根据适用性，而不是预定的来源和出处的偏好来确定数据的优先级。

我们提出的监管同行评审的证据权重方案包括以下6个通用步骤：

・定义具体监管用途及其目标，如有可能，包括具体假设的测试。

・确定不同类型数据或研究特性的加权优先次序，制定根据监管用途一般加权的参考理由。

・系统性搜索、审查和选择与新用途和假设有关的数据。

・根据其原来意图评估每项入选研究的主要、二级和三级有效性，并根据新用途评估每项研究的三级有效性。

・根据预定的算法将数据质量评估与数据加权相结合，为每项研究或数据生成一个证据权重得分。

・综合所有相关数据的证据权重得分，撰写证据权重的全面说明，叙述证据权重评估过程中得到的所有判断和结论，其中包括重要假设、不确定性以及权重因素初步形成后对其需要进行的任何调整或改进。

总结

我们回顾了背景资料，通过协调共同要素的讨论，指出正在发生趋同之处，从几方面（同行评审与GLP以及法规制定）探讨了改善和加速环境决策的基础。为实现这一目标，关键要有一个广泛适用，普遍接受的证据权重决策框架。对决策来说，需要的不是目前的个例方法，而是一个提供信息并指导各种监管决策任务的证据权重总框架。尽管有监管的具体问题，但一般证据权重原则将有利于相互沟通与提高效率。与此同时，我们注意到趋同现象，期刊同行评审正在纳入与GLP中较充分实现的概念相类似的，更高数据透明度与更多报告内容。这种趋同不仅对各种评估方案内，也对方案间评估有促进作用，并最终改进和促进同行评审、GLP应用与监管决策。

结论

监管决策中所使用科学信息的质量评估，要求对有关数据产生流程、数据汇报、分析和解释方法以及数据对上述决策活动目标的适用性做出判断。最终，这些判断需要考虑到数据的有效性和具体相关性，评估的客观性与透明度以及裁定的一致性。在有效性评估中，期刊同行评审、GLP与监管规则制定有着共同利益。然而，由于它们决策的目的不同，程序实施和所得到的结果不具备完全的可比性。

期刊同行评审在以易于理解的方式引起科学家对新的、相关而有意义数据和问题关注的过程中，实现了有价值的筛选/优化。然而，目前它并不是一个确立数据质量的可靠程序，也不是确立数据或解释以及从这些数据得出结论的相对价值的明确指标。

GLP是确立数据质量的最好方法，尤其当规定的文件要求考虑到彻底、独立再分析和重新解释时。它并非万无一失，也不提供确立解释和结论的普遍有效性或相对价值的明确指标。GLP的侧重点主要(但不完全)是确保研究遵从已认可的指导，如监管任务要求的检测具体生物反应的毒性测定。有些指导性试验经过验证程序，包括在多个实验室进行的环测试，其预测力与相关性已确定。因此，GLP代表了对用于监管决策的数据的合理选择或加权标准。

无论是同行评审还是GLP本身都不是决定数据解释以及从该解释得出结论的相对价值、普遍有效性或科学可靠性的机制。这不是它们的初衷。不存在一个用于评估广泛接受、普遍适用决策标准的金标准程序。

同行评审正在几个方面进行修订和改进。有些期刊正在应用更清晰的书面评估指导与程序。同时，很多期刊鼓励出版补充材料，提供较之主出版物上更详细的方法和结果。这些改革使期刊同行评审方法与GLP使用的方法越来越接近。

同行评审与GLP都对深入理解科学研究的结果和数据很有帮助，任何一方的缺失均不足以确立研究的相对价值和科学合理性。解决的办法是开发一个证据充分、普遍接受的证据权重方案，旨在比较、对比和评估同行评审与GLP信息以及确定相对价值和普遍有效性。该方案很容易注入监管决策程序中，利用这种数据质量评估法对具体案例进行有效决策所必需的有效性判断。

Address correspondence to L.S. McCarty, L.S. McCarty Scientific Research & Consulting, 1115 Quaker Trail, Newmarket, ON, Canada L3X 3E2. Telephone: (905) 953-9342. E-mail:lsmccarty@rogers.com

Financial support was provided by CropLife America. L.S.M. is employed by L.S. McCarty Scientific Research & Consulting; C.J.B. is employed by Applied Pharmacology and Toxicology, Inc., Gainesville, Florida, USA; and E.M.M. is employed by Environmental and Regulatory Resources, Durham, North Carolina, USA.

The conclusions, opinions, comments, and positions expressed herein are solely the authors' and may not reflect those of CropLife America, its employees, or member companies. There are no contractual relations or proprietary considerations that restrict the authors' publication or dissemination of findings.

Received 29 July 2011; accepted 17 February 2012.

译自*EHP* 120(7):927−934(2012)

翻译：王仁礼；审校：李卫华

★本文参考文献请浏览英文原文

原文链接

http://dx.doi.org/10.1289/ehp.11042773