



建立科学透明度机制

David Mellor 博士 (ORCID 0000-0002-3125-5888)

开放科学中心项目经理

david@cos.io | @EvoMellor | <https://osf.io/qthsf>

莫顿 (Robert Merton) 提出了四项著名的现代科学的惯性做法¹：通用性，共用性，无私和有组织的怀疑主义。这些惯性做法的核心是基于科学真理不是由某人的名声或由于支持某理论的欲望去决定，相反的，证据是由一个社群的广大共识去定论。尽管这些规律都为科学社群所广泛接受，有些研究人员承认他们在实际操作上未能达至这些理想，大部分研究人员也觉得他们的同行没有落实这些规律。²身处这样的环境里，作出改变对任何个人来说都是有风险的。

虽然研究人员都相信理想的科学规条，但他们不能从理想主义中获益，他们只能通过他们的文章发表数得到嘉许，而这些著作都要求创新和突破性的结果。³就导致一个现象，发表的学术文章都偏向统计学上有显著差异的结果。⁴这表示发表的学术文章并不真实反映我们知识的全部。

出版正面、创新结果让作者有很强的诱因在没有意识下出版有偏差的研究结果。此外，有很多方法可以让研究人员扭曲结果，让结果看起来更创新和鼓舞人心。例如 HARKing (hypothesizing after results are known)，意即在分析结果后再作出假设，就是其中一种把事后分析结果汇报成好像是事前⁵计划一样。同样地，“p-hacking”是在做了很多统计测试以后，选择只汇报把一小簇统计测试的过程，这通常是很短的过程并没有打算扭曲整体证据的意图。对统计分析细微和看起来合理的改变可以很容易将确认性假设测试转变成脆弱的资料库探索。而这个做法很常见。⁶

John Ioannidis 预测理想科学和实际做法之间的冲突会导致研究结果不能重复。⁷开放科学合作联盟 (The Open Science Collaboration) 尝试复制研究结果，但最后只有不到一半的研究能够重复。⁸改变科学生态系统去认可理想的实践行动似乎是不容置疑的目标。

1 Merton RK. "The Normative Structure of Science." In *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: The University of Chicago Press; 1973:267–278

2 Anderson MS, Martinson BC, De Vries R. Normative dissonance in science: results from a national survey of U.S. scientists. *J Empir Res Hum Res Ethics*. 2007;2(4):3–14. [Crossref](#)

3 Nosek BA, Spies JR, Motyl M. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspect Psychol Sci*. 2012;7(6):615–631. [Crossref](#)

4 Fanelli D. "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*. 2010;5(4):e10068. [Crossref](#)

5 Kerr NL. HARKing: Hypothesizing after the results are known. *Pers Soc Psychol Rev*. 1998;2(3):196–217. [Crossref](#)

6 Head ML, Holman L, Lanfear R, et al. The extent and consequences of p-hacking in science. *PLoS Biol*. 2015;13(3):e1002106. [Crossref](#)

7 Ioannidis JPA. Why Most Published Research Findings Are False. *PLoS Med*. 2005;2(8):e124. [Crossref](#)

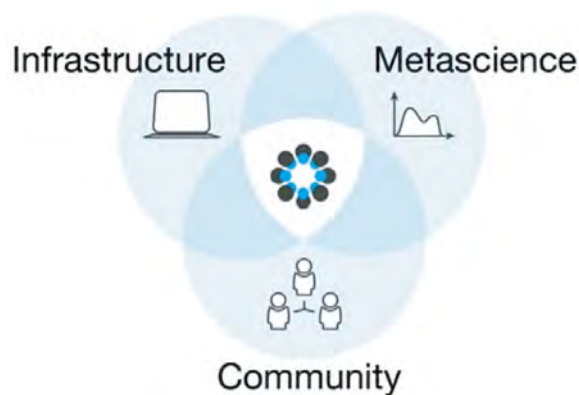
8 Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015;349(6251):4716–4716. [Crossref](#)

开放科学中心 (Center for Open Science, COS) 在 2013 年成立, 目的是增加科学的重复性。想要增加科学研究的重复性就要增加科学工作流程的透明度, 因为能让专家审查收集证据时的每一个重要步骤。如果研究工作流程的每一方面都可以重复互相连接起来, 我们可以减少重复前人研究的障碍, 加强批判审查研究的能力。要达到这个远景并不容易。研究人员因着面前的奖励机制而导致当下的情况, 而我们必须正视和改变这些机制。这需要整个学术生态系统的集体行动去奖励更优秀的做法: 出版商、大学、学术组织和资助机构。

正因为这项挑战, COS 集中精力在三项主要活动。首先, 我们的元科学团队会评估问题的严重性并提出解决方法。我们在心理学⁷和癌症生物学的领域中进行研究重复性的计划, 藉此估计在一个学科中重复研究结果的可能性。我们在支援 [Many Lab](#) 各项计划中尝试重复不同环境下单项研究的结果, 并估算每项结果的边界限制。我们也审查了我们的倡议的有效性, 例如“开放实践徽章” (Open Practice Badges) 对增加分享数据和研究材料的效果。⁹

其次, 我们通过社群团队从事教育和宣导工作, 希望鼓励更好的研究实践。这不限于举办[讲座](#)、[网络研讨会](#)和制作有关如何重复研究的[资源](#), 也包括制作政策指引和举行教育活动去鼓励大家采用此做法。我们的宣导工作希望可以改变现有研究生态系统中的诱因, 回报精确和透明的研究实践, 而不是哗众取宠但不能重复的结果。

9 Kidwell MC, Lazarevic LB, Baranski E, et al. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biol.* 2016;14(5):e1002456. [Crossref](#)



图一: 我们的元科学、社群和基础团队一起合作去评估问题的严重性, 提倡和教育改变, 并建立需要的工具去落实解决方法。

我们的政策推广工作中的一个例子是透明和开放推广基础元科学指引 (Transparency and Openness Promotion (TOP) Infrastructure Metascience Guidelines)。TOP 指引提供了八个单元式标准, 学术期刊、出版商和资助机构都可以采用这个指引来奖励透明胜而不是创新。这些标准包含了不同的做法, 例如数据引用、数据和材料共享、设计标准、预先注册和重复研究。

为了减少采用 TOP 指引的障碍, 八个标准的每一个都可以在三级中的任何一级被采用以增加严谨度。第一级要求研究人员对他们的行动作出声明, 第二级要求行动, 第三级则需要确认行动。例如, 第一级中可以采用数据透明性的标准, 这项标准要求研究人员**声明**资料是否会公开让公众取得。第二级**要求**数据共享 (除了编辑部有伦理考量例外)。第三级包括**确认**数据可以用在复制研究的主要结果。分级设计可以减低采用指引的困难, 也提供改善未来工作的指引。虽然很少研究人员会打算采用最严谨的标准, 公开和分享数据的规定可以随时推行。

预先注册挑战 (Preregistration Challenge) 是一个比赛, 一千名研究人员可以透过发表他们预先注册的成果获得一千



图二：预先注册可以确保决定是在看到数据前做出的。发表预先注册结果的研究人员可以成为千人之一的得奖者。

美元的奖金。教育活动的设计是希望鼓励采用预先注册，重要的分析决定可以在进行研究前就已注明。预先注册清楚的分开了假设检验（确认性）的研究和产生假设（探索性）的研究。

当研究人员在数据集中进行探索时，这两个步骤的差别会变得模糊。每一个在分析时做的决定实际上可以衍生出一个新的假设。例如，每一个关于剔除法则、停止规则、把量度的变数并入成果指数还有控制变数等的决定，都有机会形成一个独立的分析。由于这些分析的单一子集都有机会在统计学上达至显著的差异，加上研究人员有很强的诱因去找出和发表单一子集的，不管整体证据为何，很多结果都可以呈现成统计显著的结果。¹⁰

一个关于预先注册的忧虑是会给研究人员太多束缚。假如失去发现意外结果的弹性，我们可能会失去什么？科学诚然需要偶然的发现，且有机会错过意外的发现。第二型误差 (Type II error) 和假阴性 (false negatives) 会阻碍我们的进度。可是当我们要确认一个意外的发现，我们应该保留 P 值的功用并减少假阴性。简单的把产生假设与假设检验的工作分清楚就已经可以

10 Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci. 2011;22(11):1359–1366. [Crossref](#)



图二：开放科学架构把整个研究流程连接起来，解决了大部分研究实验室常见的问题：合作、移交和坚持。

增加透明度和结果的肯定性。预先注册排除了误把分析呈现成确认分析的诱因，让研究人员有机会探索数据集，找出预期外的结果。

最后，我们的工作的第三个重点是创造一个可以做到我们提倡的改变的基础。开放科学架构 (Open Science Framework, OSF) 是我们的旗舰产品，它是一个免费的开放源代码工作流程管理工具，可以帮助指导透明和可重复的工作。OSF 能让研究人员管理复杂的研究计划和跟同行的合作。这个工具有内置功能鼓励用户以独特、一致的标记分享数据。用户可以用这个工具在不同时间点注册研究计划。这些研究计划注册是只读记录，一个研究人员的计划的一致「快照」。研究计划可以在研究周期的不同时间点注册，借此保存状况。预先注册只是其中一种快照，是在收集数据前截取的。其他完成注册截取的时间点还有投稿前和文章发表前。

OSF 可以作为一个院校中进行研究的枢纽。跟传统方法相比（请参考弗吉尼亚大学的 OSF 网页作例子）这些院校的网页主页可以协助浏览者更快速和容易找到合作伙伴和现在正在进行的研究。

改善已发表科学文献的重复性是一个复杂的挑战，这个难题只能通过社群去解决。COS 鼓励通过推广和已建立的工具去完成改变。

你如何可以参与这项改革？期刊编辑可以透过成为 TOP 指引的连署人去推广这些价值。编辑也可以颁发[开放实践徽章](#)或在得知结果前利用[注册报告](#)的范例进行同行评审。研究人员可以参加[预先注册挑战](#)

厘清在看到数据前或后决定分析方法的分别。他们可以用 [OSF](#) 管理研究工作，与伙伴的合作或分享数据。

我们的远景是所有在研究领域中的工作流程都是透明的。假如有需要，这种透明性可以通过专家评估改善严谨性。缺少集体行动就无法达到这个远景，因此请加入我们，改善研究的生态系统。