

化学工程研究的再现性问题

大卫·沃尔顿, 丽贝卡·肖尔, 克里斯塔·汉

隶属机构: 美国化学与生物分子工程学院

摘要: 近年来, 多个科学领域都对已发表结果的可再现性提出了担忧。在生物医学和心理学领域已经进行了系统性的研究来探讨这个问题, 但在支撑现代化学工程的材料导向研究中, 人们对这个重要问题知之甚少。在这里, 我们讲述了我们自己机构的一个戏剧性历史事件, 以说明进行可再现研究的含义, 并描述了两个基于文献分析的案例研究, 以提供有关现代材料研究的可再现性的具体信息。这两个案例研究涉及金属有机框架(MOF)的特性, 这是一类已产生数万篇论文的材料。我们并不声称对MOF的研究比其他子领域的可再现性更低(或更高); 相反, 我们认为这个子领域的特征在面向材料的研究的许多领域中是共同的。

关键词: 再现性; 材料化学; 吸附; 晶体结构; 金属有机骨架

The Reproducibility Problem of Chemical Engineering Research

David Walton, Rebecca Sholl, Krista Han

Affiliation: School of Chemical and Biomolecular Engineering, USA

Abstract: Concerns have been raised in multiple scientific fields in recent years about the reproducibility of published results. Systematic efforts to examine this issue have been undertaken in biomedicine and psychology, but less is known about this important issue in the materials-oriented research that underpins much of modern chemical engineering. Here, we relate a dramatic historical episode from our own institution to illustrate the implications of performing reproducible research and describe two case studies based on literature analysis to provide concrete information on the reproducibility of modern materials-oriented research. The two case studies deal with the properties of metal-organic frameworks (MOFs), a class of materials that have generated tens of thousands of papers. We do not claim that research on MOFs is less (or more) reproducible than other subfields; rather, we argue that the characteristics of this subfield are common to many areas of materials-oriented research. We conclude with specific recommendations for action by individual researchers, journal editors, publishers, and research communities.

Keywords: Reproducibility; materials chemistry; adsorption; crystal structures; metal organic frameworks

引言:

尽管科学哲学家可能会争论科学方法的原则, 但将化学工程等应用学科的核心目标描述为生产具有可靠性能和规格的技术和产品是没有争议的。关注可靠性的一个实际原因是关于财务方面的: 建设大型化工厂需要大量的前期投资。这意味着对于这样一个工厂内的过程是否会起作用, 可以容忍的风险水平非常低, 也许几乎为零。由于化学工艺设计通常依赖于指定各种材料的物理特性, 因此必须可靠地了解这些特性, 然后才能做出明智的设计决策。全球有大量研究小组开展研究, 重点是开发新材料(例如催化剂、吸附剂、可调溶剂、电池阴极、太阳能电池), 目的是让这些材料用于现实世界的工

程应用。本文的目的是探讨一个重要的(即使是令人不安的)问题: 化学工程和相关领域中以材料为导向的研究结果的可重复性如何?

几乎每个活跃的研究人员都可以描述以前无法复制的工作示例。在许多情况下, 这些战争故事是大量时间和金钱浪费在关注文献中先前工作的结果。更坦率的研究人员通常可以找出即使在他们的团队内部也无法重复他们自己团队中以前的工作的例子。近年来, 人们对这些问题的广泛性引起了对多个科学领域的强烈兴趣, 包括将科学描述为处于危机中的书籍长度的处理方法。

Ioannidis 于 2005 年所发表的一篇命题为“为什么大多数已发表的研究结果都是错误的”的文章激起了人们对科

学可重复性的担忧。这篇文章使用了一个简单的统计模型来得出标题中描述的结论。然后,约阿尼迪斯从他的结论中得出了几个推论。前两个符合科学常识:小型研究更容易产生不正确的发现,而小的物理效应更容易产生不真实的结论。然而,其他推论涉及更多的社会评论,例如,“科学领域越热,研究结果就越不可能是真实的”。

几项开创性研究以最直接的方式解决了可重复性问题:通过重复实验。在心理学方面已经做出了广泛的努力,部分原因是未能复制诸如力量姿势之类的效果,这些效果在首次报道时引起了公众的极大关注。例如,一个大型国际心理学家团队在36个独立样本中重复了13个“经典和现代效应”。他们发现13种效应中有10种复制一致,但其余3种要么没有复制,要么复制很弱。该领域的研究问题和方法与面向材料的工程研究有很大不同,很难从另一个方面得出结论。一项与化学工程师产生更强烈共鸣的引人注目的研究描述了拜耳多年来的工作,旨在复制已发表的肿瘤学和药物方面的学术发现。这些项目通常持续6-12个月。在报告的67个项目中,仅21%实现了完全复制,而在65%中观察到“不一致”。一年后,来自Amgen的一个团队报告了类似的结果,他们协调努力复制了53篇血液学和肿瘤学领域有史以来被引用次数最多的论文。在这种情况下,原始论文的主要科学发现只有11%的时间得到证实。思考这一结果可以很容易地理解科学处于危机中的叙述是如何产生的,尽管有些人反对这种描述。

上述拜耳和安进研究的结果令人震惊,但这些发现能否外推到面向材料的工程研究?毕竟,肿瘤学和类似领域的生命科学研究本质上依赖于生物样本,许多人认为,至少在传闻中,这些样本比非生物材料更具可变性。我们在本文中的重点是考虑在非生物领域的材料导向研究中的可重复性研究(为简洁起见,我们在下面简称为材料导向研究)。描述这些领域典型研究问题的一些特征是有用的。首先,这些问题的特点是有许多(可能是103-106种)不同的材料或材料成分,可以在寻找获胜材料时加以考虑。热门话题的研究文献通常包含许多描述密切相关材料的合成和测试的论文。这些问题的第二个共同特征是研究界使用定义明确的指标来判断哪种材料是最好的。这些指标的三个例子是太阳能电池的转换效率、热电材料的品质因数(ZT)以及用于气体捕获的吸附剂的存储容量。在这三个例子中,太阳能电池因其广泛的商业应用而脱颖而出。有趣的是,太阳能电池与更多研究型材料的不同之处在于,有一种被广泛接受的第三方机制来建立光伏效率性能测量的一致性和准确性。

尽管科学文献中的大量努力通常集中在一个简单的性能指标上,但重要的是要记住,为实际应用选择材料通常受多种标准的影响。这些领域的第三个特点是,成功不仅带来科学荣誉,还带来知识产权和商业价值的潜在利益。这第三个特征可能会在数据的公正和及时共享方面造成相当大的复杂性。

案例研究一:吸附金属有机框架中的二氧化碳

在这里,我们总结了Park等人最近的工作,他们采用了另一种方法,即在现有文献中搜索已经存在的重复实验。这种文献荟萃分析策略利用了热门研究领域的一个特点:让许多小组研究密切相关的材料通常会导致重复实验,即使重复不是所做工作的具体目标。

Park等人专注于MOF中的CO₂吸附。MOF是一类结晶纳米多孔材料,在过去15年中引起了广泛的兴趣,用于各种应用,包括气体吸附。以前没有接触过这些材料的读者可以获得对MOF的物理性质和合成的优秀评论。已经发表了数万篇与MOF的合成及其性能相关的论文。由于全球对CO₂捕获作为缓解气候变化的可能工具的浓厚兴趣,因此有大量论文可用于测量MOF中CO₂的吸附特性。尽管基于吸附的完整分离系统的设计涉及许多因素,但CO₂的平衡吸附等温线经常被用来判断MOF(或其他吸附剂)是否有用或有趣。Park和他的同事讨论的主题是可以明确说明MOF中CO₂吸附等温线的再现性。这个主题的选择是基于对特定物理现象的实验再现性做出坚定陈述的可能性,而不是因为Park等人。相信MOF中的CO₂吸附在某种程度上比在面向材料的研究中广泛研究的其他物理现象更具(或更少)可重复性。

Park等人研究的一个关键方面是CO₂吸附等温线的实验测量是相对常规的,可以使用几种广泛可用的商业仪器进行。美国国家标准与技术研究院(NIST)最近的一项研究有力地说明了这一点,其中11个独立小组测量了NIST提供给每个小组的沸石在20°C时的CO₂吸附量。这项工作的13项实验的结果是使用一系列设备进行的。实验之间的良好一致性是显而易见的。这项研究并不意味着进行这些测量是微不足道的——仍然需要小心和技巧才能准确地进行测量。尽管如此,结果确实表明,至少在原则上,在室温下可靠测量多孔材料中的CO₂吸附对于装备精良的研究人员来说并不构成巨大挑战。

Park等人工作的第二个关键方面。是存在一个公开可用的数据库,该数据库试图以全面的方式从公开文献中收集平衡吸附等温线(对于所有分子,而不仅仅是CO₂)。收集这些数据需要NIST领导的编译团队付出巨大的持续努力,而没有这些数据需要Park等人的工作。范

围会更有限。

Park 等人分析了 NIST 吸附数据库中报告的数千条等温线，以收集 MOF 中 CO₂ 的所有可用测量值。从这些数据中，他们确定了多次报告的所有示例。对于少量材料，可以使用相对大量的复制品。重复次数最多的示例是著名的 MOFHKUST-1，显示了该材料在小温度范围内可用的所有 18 条等温线。众所周知，HKUST-1 对水分敏感，因此材料制备和处理的微小变化可能导致材料性能发生重大变化。

有两个强烈对比的重复测量纳米多孔材料中 CO₂ 吸附的示例，其中一个具有高度可重复性，而另一个则不是。哪个例子更有代表性？Park 等人的工作对这个问题提供了一些见解，因为他们将上述方法扩展到所有报告的 MOF 中的 CO₂ 等温线。他们的主要发现之一是，在分析的所有重复等温线中，五分之一（或更准确地说，21%）被归类为异常值。也就是说，五分之一的等温线在统计上与相同材料的其他可用复制品不一致。这是一个发人深省的观察，因为如果它代表了整个文献，这意味着从文献中选择一条等温线大约有五分之一的机会使数据与通常被认为是正确的数据有很大差异回答。对于其他面向材料的研究领域来说，确定这一估计是否也是一个合理的估计显然是非常有趣的。

如果可以访问足够多的重复测量值，可以很容易地检测到异常值，那么上面讨论的 20% 的异常值出现就不会令人担忧。不幸的是，这并没有描述与面向材料的研究相关的文献的现实。尽管已经发表了数千篇关于 MOF 中气体吸附数据的论文，但 Park 等人发现，在去除异常值后，只有九种材料存在四个或更多复制 CO₂ 等温线。对于公开文献中的绝大多数 CO₂ 吸附等温线，没有可用的独立复制品。

我们强调 Park 等人的结果集中在 MOF 中的气体吸附，因为可以获得全面的数据收集，而不是因为作者认为使用这些材料的实验本质上比任何其他领域的测量具有更多（或更少）可重复性以材料为导向的研究。我们对其结果的总结应仅被视为研究领域可重复性的初步案例研究。关于从此类回顾性文献分析中可以推断出什么和不能推断出什么，存在一系列问题，我们建议感兴趣的读者参考 Park 等人的原创作品以了解更多细节。

案例研究二：晶体化纳米多孔材料的复制合成

第一个案例研究侧重于纳米多孔材料的特定应用，即二氧化碳的吸附。此特定应用可能会低估有关基础材料可重复性的已知信息，因为后续工作可能会遵循最初报告的材料合成，但会针对不同应用测试此材料。这提

出了一个更基本的问题，可以询问有关面向材料的研究的可重复性：一旦报告了新材料的合成，该合成多久独立复制一次？如果文献中的新材料不再制造（或者更准确地说，公开文献中不再报道它们的合成），那么不可避免地会导致对这些材料的再现性或它们的特性一无所知。

为了探讨我们上面提出的问题的一个具体例子，我们检查了描述 MOF 材料合成的文献。关于 MOF 及其特性的论文已经发表了数万篇，因此很难说社区没有足够的机会重复材料的合成。尽管尝试对这些文献进行全面分析会很有趣，但这样做远远超出了我们在此讨论的范围。相反，我们选择了一小部分材料并检查了它们在文献中的历史。因此，我们在本节中的讨论应被视为轶事，而不是提供明确的分析。

新晶体材料（如 MOF）的晶体结构通常在出版期间或之后在剑桥结构数据库（CSD）中报告。我们从 Computation-Ready, Experimental (CoRE) MOF 数据库中选择了 16 个 MOF，该数据库是 CSD 的一个子集，几乎包含所有无序 MOF。这 16 种材料中的每一种最初都是在 2007 年至 2013 年期间报告的，并且在可能的情况下，我们选择了原始报告描述单一材料的示例。后一种选择大大简化了对后来引用原始结果的论文的分析。在撰写本文时，这 16 篇原始论文被引用了 8 到 168 次，平均被引用次数约为 50 次。当然，可能有正当理由不重复对特定材料的研究；例如，材料在首次针对特定任务进行测试时的性能可能很差。如果一篇论文被多次引用，这至少是一个微弱的证据，表明所报道的材料具有足够的价值，让社区对其感兴趣。

我们检查了引用原始 16 份合成报告之一的 768 篇论文中的每一篇，以确定哪些材料已被再次合成。我们没有对重复合成的质量或结果做出判断；如果作者说他们制作了与早期论文相同的材料，那么我们认为这是重复合成。我们将具有相同名称的 MOF 的化学变体，例如 MIL-96 (Al) 和 MIL-96 (Cr) 视为复制合成。如果有机接头被官能化，如果被取代的金属中心从根本上改变了感兴趣的主要性质（例如，晶体结构），或者如果客体夹杂物显著改变了感兴趣的主要性质，则认为引用进行了修改合成。我们下面的讨论不包括重复计算；如果一篇论文重复了精确合成然后进一步修改它，我们将其视为精确重复合成的一个例子，但不包括修改后的合成。我们还检查了施引论文，看它们是否与原始报告共享一位或多位作者。

在我们调查的 16 种材料中，超过四分之一（5/16）没有被任何作者再次合成，超过一半（9/16）没有由独

立于原作者的一组作者合成.换句话说,从已发表的文献中无法得出任何关于大多数材料的再现性的结论,即使是由两个独立小组制作的材料的最低水平。只有三种材料的合成精确重复了两次以上,而这些材料中只有一种是由独立于原作者的研究人员进行的这些复制。有趣的是,报告的重复次数与初始论文的引用次数并没有很强的相关性。例如,原始作者复制次数最多的材料仅是我们检查的论文中被引用次数第五多的论文。

也许我们的结果中最引人注目的观察是,报告的修改合成的数量明显大于精确合成的数量。回想一下,在我们的符号中,仅当该报告不包括原始材料合成的数据时,才将材料记录为具有修改的合成报告。似乎在许多开发改进合成的研究中,原始材料也是在新工作的过程中产生的。这表明作者在报告他们的工作时有意识地选择不包括有关复制合成的信息。如果这是正确的,那么研究界就错失了通过报告重复实验来增强已发表文献的可靠性的机会。

我们提供的轶事数据虽然发人深省,但应该被认为是说明性的,而不是确定性的。我们只检查了以非随机方式选择的少数示例。我们假设这些合成MOF的结果表明将相同的方法应用于其他类别的材料会发现什么,但我们尚未验证这一假设。我们的分析只检查材料是否被重复合成,而不是重复合成是否给出一致结果的问题。后一个问题对于MOF合成的定量解决具有挑战性,因为评估合成结果的主要工具粉末X射线衍射通常是定性而非定量工具。

结论和建议

前面的部分旨在促使读者思考他们自己研究领域的可再现性状态。在理想主义的层面上,很容易得出结论,如果文献中信息的可靠性更高,或者更容易评估,那么以材料为导向的研究对社会的长期影响和价值会得到改善。同时,任何提高可重复性的努力都伴随着相对于当前规范的实际时间和资源成本。对科学可复制性的担忧已导致研究人员个人以及机构、期刊和资助机构在社区层面提出一系列改进建议。尽管其中许多建议具有广泛的适用性,但生命科学(例如生物医学)、行为科学(例如心理学)和面向材料的化学工程之间的研究文化差异意味着在前几个领域提出的建议并不总是能很好地转化进入后者。例如,在生物医学和行为科学中,一项重要的建议是增加研究的样本量,以增强可获取信息的统计能力。得知基于材料的工程论文中报告的典型测量值 $N=1$,这些领域的许多科学家会感到震惊。也就是说,报告来自单个样本或实验的数据是很常见的。对这

一认识的适当回应是旨在量化关键测量中的不确定性来源(通过误差条或其他方法)。这样说来,这个建议在极端情况下似乎没有争议,但本文的资深作者可以很容易地在我们已发表的作品中没有找到明确报告这项工作的例子。考虑一下您自己的研究小组发表的作品是否也是如此,这将是一个有用的练习。

参考文献:

- [1]Garrison KE, Tang D, Schmeichel BJ. 2016. Embodying power: a preregistered replication and extension of the power pose effect. *Soc. Psychol. Personal. Sci.* 7:623 - 30.
- [2]Klein RA, Ratliff KA, Vianello M, Adams R, Bahn í k Š, et al. 2014. Investigating variation in replicability: a “many labs” replication project. *Soc. Psychol.* 45:142 - 52.
- [3]Prinz F, Schlange T, Asadullah K. 2011. Believe it or not: How much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10:712.
- [4]Begley CG, Ellis LM. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* 483:531.
- [5]Fanelli D. 2018. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *PNAS* 115:2628 - 31.
- [6]Walton KS, Sholl DS. 2017. Research challenges in avoiding “showstoppers” in developing materials for large-scale energy applications. *Joule* 1:208 - 11.
- [7]Park J, Howe JD, Sholl DS. 2017. How reproducible are isotherm measurements in metal-organic frame- works? *Chem. Mater.* 29:10487 - 95.
- [8]Eddaoudi M, Kim J, Rosi N, Vodak D, Wachter J, et al. 2002. Systematic design of pore size and func- tionality in isoreticular MOFs and their application in methane storage. *Science* 295:469.
- [9]Lee J, Farha OK, Roberts J, Scheidt KA, Nguyen ST, Hupp JT. 2009. Metal-organic framework materials as catalysts. *Chem. Soc. Rev.* 38:1450 - 59.
- [10]Cohen SM. 2012. Postsynthetic methods for the functionalization of metal-organic frameworks. *Chem. Rev.* 112:970 - 1000.
- [11]Keskin S, van Heest TM, Sholl DS. 2010. Can metal-organic framework materials play a useful role in large-scale carbon dioxide separations? *ChemSusChem* 3:879 - 91.
- [12]Maring BJ, Webley PA. 2013. A new simplified pressure/vacuum swing adsorption model for rapid ad- sorbent screening for CO₂ capture applications. *Int. J. Greenh. Gas Control* 15:16 - 31.

