Annual Review of Chemical and Biomolecular Engineering

Does Chemical Engineering Research Have a Reproducibility Problem?

Rebecca Han, Krista S. Walton, and David S. Sholl

School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332–0100, USA; email: david.sholl@chbe.gatech.edu

Annu. Rev. Chem. Biomol. Eng. 2019. 10:43–57

First published as a Review in Advance on March 27, 2019

The *Annual Review of Chemical and Biomolecular Engineering* is online at chembioeng.annualreviews.org

https://doi.org/10.1146/annurev-chembioeng-060718-030323

Keywords

reproducibility, materials chemistry, adsorption, crystal structures, metal-organic frameworks

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.

1. INTRODUCTION

Although philosophers of science may debate the tenets of the scientific method, it is not controversial to describe a core aim of applied disciplines such as chemical engineering as producing technology and products with reliable performance and specifications. One pragmatic reason for this focus on reliability is financial: Constructing a large-scale chemical plant requires very large up-front investments. This implies that the level of risk that can be tolerated about whether the processes within such a plant will work is very, perhaps vanishingly, low. Because chemical process designs typically rely on specifying physical properties of various materials, it follows that these properties must be known reliably before sensible design decisions can be made. An enormous number of research groups around the globe perform research focused on developing new materials (e.g., catalysts, adsorbents, tunable solvents, battery cathodes, solar cells) with the aim of having these materials used in real-world engineering applications. The aim of this article is to examine an important, if uncomfortable, question: How reproducible are research results in materials-oriented research in chemical engineering and allied fields?

Almost every active researcher can describe examples of previous work that could not be reproduced. In many cases, these war stories are the outcome of a considerable waste of time and money devoted to following previous work in the literature. More candid researchers can often identify examples where previous work from within their own group could not be repeated even within their group. Concern about how widespread these issues are has generated intense interest in multiple fields of science in recent years, including book-length treatments that have characterized science as being in crisis (1).

Concern about reproducibility in science was galvanized by a 2005 article by Ioannidis (2), provocatively titled, "Why Most Published Research Findings Are False." The article used a simple statistical model to draw the conclusion described in its title. Ioannidis then drew several corollaries from his conclusion. The first two are consistent with scientific common sense: Small studies are more likely to generate incorrect findings, and small physical effects are more likely to generate untrue conclusions. Other corollaries, however, involve more social commentary, for example, that "the hotter a scientific field…the less likely the research findings are to be true" (2). We relate a particularly striking example of this phenomenon in Section 2.

Several seminal studies have tackled the issue of reproducibility in the most direct way possible: by repeating experiments. Extensive efforts of this kind have been made in psychology, spurred in part by failures to replicate effects, such as power poses, that had drawn enormous public attention when they were first reported (3). For example, a large international team of psychologists repeated 13 "classic and contemporary effects" across 36 independent samples (4). They found that 10 of the 13 effects replicated consistently, but the remaining three either were not replicated or were replicated weakly. The research questions and methods in this field differ so much from materials-oriented engineering research that it is difficult to draw conclusions about one from the other. A striking study that resonates more strongly with chemical engineers described years of work at Bayer aimed at reproducing published academic findings in oncology and other aspects of drug discovery (5). These projects typically lasted 6–12 months. In the 67 projects reported, complete replication was achieved in only 21%, whereas "inconsistencies" were observed in 65%. Similar results were reported a year later by a team from Amgen, who coordinated an effort to reproduce 53 of the most highly cited papers of all time in hematology and oncology (6). In this case, the major scientific findings of the original papers were confirmed only 11% of the time. Pondering this outcome makes it easy to understand how a narrative of science being in crisis can arise [although some have pushed back against this characterization (7)].

The outcomes from the Bayer and Amgen studies described above are striking, but can these findings be extrapolated to materials-oriented engineering research? After all, life-sciences research in oncology and similar fields inherently relies on biological samples, which many would view, at least anecdotally, as more variable than nonbiological materials. Our focus in this article is to consider how reproducible research is in materials-oriented research in nonbiological fields (which we refer to simply as materials-oriented research below for brevity). It is useful to describe some characteristics of typical research problems in these areas. First, these problems are characterized by having many, perhaps 10^3–10^6, different materials or material compositions that can be considered in a search for the winning material. The research literature in a hot topic (to use Ioannidis's term) often contains many papers describing synthesis and testing of closely related materials. A second common characteristic of these problems is that the research community uses well-defined metrics to judge which material is the best. Three examples of these metrics are the conversion efficiency of solar cells, the figure of merit (ZT) for thermoelectric materials, and the storage capacity of adsorbents for gas capture. Among these three examples, solar cells stand out because of their widespread commercial deployment. It is interesting to note that solar cells differ from many more research-oriented materials in that there is a widely accepted third-party mechanism for establishing consistency and accuracy among performance measurement of photovoltaic efficiency (8). Although considerable effort in the scientific literature often focuses on a simple performance metric, it is important to remember that selection of a material for a practical application is typically influenced by multiple criteria (9). A third characteristic in these areas is that success brings not only scientific accolades but potential benefits from intellectual property and commercial value. This third characteristic can create considerable complications in unbiased and timely sharing of data.

The remainder of this article is organized as follows. Section 2 recounts a dramatic historical example of the consequences of reporting data that are not reproducible. Sections 3 and 4 describe two specific efforts to assess the state of reproducibility in a focused area of materials research, the synthesis and use of metal-organic frameworks (MOFs) as adsorbents. Work on MOFs has all of the characteristics described above, but the choice of this topic area is not intended to imply that extant work on MOFs is more (or less) reproducible than other similar topics that could be studied. Section 5 is perhaps the most important; in it we give recommendations for individual researchers and research communities to improve reproducibility in materials research.

2. A CAUTIONARY TALE: COLD FUSION AT GEORGIA TECH

Before discussing the broader issues associated with experimental reproducibility, it is useful to pause and consider the possible consequences of being involved in work that cannot be reproduced. We do this by relating an extreme example from our own institution. This account is drawn primarily from the detailed (and highly entertaining) account by one of the lead scientists, James Mahaffey (10), but also from newspaper archives.

On March 23, 1989, Stanley Pons and Martin Fleischmann stunned the scientific world by holding a news conference describing their discovery of room-temperature (cold) fusion. In an astonishing coincidence, the Exxon Valdez ran aground in Alaska the next day, causing an enormous oil spill and reinforcing societal desire for clean energy sources. A scientific gold rush ensued, with research teams around the world scrambling to build and test their own cold fusion devices using information gathered from the Pons and Fleischmann news conference and other media appearances. Six days after the initial news conference, several researchers at the Georgia Tech Research Institute led by James Mahaffey submitted an internal proposal for $25,000 to perform cold fusion experiments. The funds were granted within hours, an outcome that many readers can only dream about for their own grant applications.

After four more days, the Georgia Tech team had built their apparatus and was performing experiments. After two days of experiments, the team repeatedly detected neutrons being generated, a critical sign of fusion that was a key element of confusion in Pons and Fleischmann's original experiments. In short order, a press conference was convened that announced the Georgia Tech findings, an event that attracted national press attention. A front-page story in the *Atlanta Constitution* the next day was headlined, "Ga. Tech Reports Cold Fusion Test Apparently Backs Controversial Finding," and went on to say, "Georgia Tech researchers…have attained controlled, room temperature nuclear fusion."

The researchers' elation was short-lived. Further tests raised serious concerns about their initial interpretation, ultimately leading to the principled decision to hold another press conference three days after their first one. The next day the story was again front-page news in the *Atlanta Constitution*: "Tech Scientists Retract Fusion Claim, Must Repeat Experiment." The *New York Times* headline was stronger: "Georgia Tech Team Reports Flaw in Critical Experiment on Fusion." The team had discovered that their neutron detector was highly temperature sensitive, and what was originally reported as excess neutrons was simply background counts associated with the detector heating up in the experimental apparatus. In time, they realized what many other efforts to repeat the original cold fusion experiments also showed: The process did not generate excess neutrons or excess heat. Cold fusion was a flop.

Many useful lessons can be drawn from this piece of scientific history. After marveling at the craziness associated with this episode, it is worth thinking seriously about its implications for your own research. Aside from the obvious lesson in the extreme peril in performing science by press conference, think about how the individuals involved felt when their work was found to be false. How would you feel if the research project you had diligently worked on for months or years was repeated by others and their conclusions contradicted or undermined your key conclusions? How would you feel if, when your research was published, you had a nagging feeling that others would not be able to reproduce it? Serious reflection on these kinds of questions is a valuable exercise and provides strong motivation to understand whether reproducibility challenges exist in one's own research area and how research can be performed and reported to reduce these challenges.

3. CASE STUDY 1: CO_2 ADSORPTION IN METAL-ORGANIC FRAMEWORKS

After the historical diversion of the previous section, we now return to considering how reproducible materials-oriented experiments are. One way to examine this issue, of course, is to systematically repeat experiments from the literature in a controlled way. Doing so, however, requires a significant investment of time and resources. Here, we summarize recent work by Park et al. (11), who took the alternative approach of searching the existing literature for replicate experiments that already exist. This literature meta-analysis strategy takes advantage of a characteristic in hot areas of research: Having many groups work on closely related materials can often lead to replicate experiments, even if replication is not the specific aim of the work being done.

Park et al. (11) focused on CO_2 adsorption in MOFs. MOFs are a class of crystalline nanoporous materials that have drawn intense interest over the past 15 years for a variety of applications, including adsorption of gases. Excellent reviews of the physical properties and synthesis of MOFs are available for readers who have not previously encountered these materials (12–14). Tens of thousands of papers have been published associated with the synthesis of MOFs and their performance. Because of strong worldwide interest in CO_2 capture as a possible tool to mitigate climate change, an enormous number of papers are available that measure the adsorption properties of CO_2 in MOFs (15). Although design of a complete adsorption-based separation system

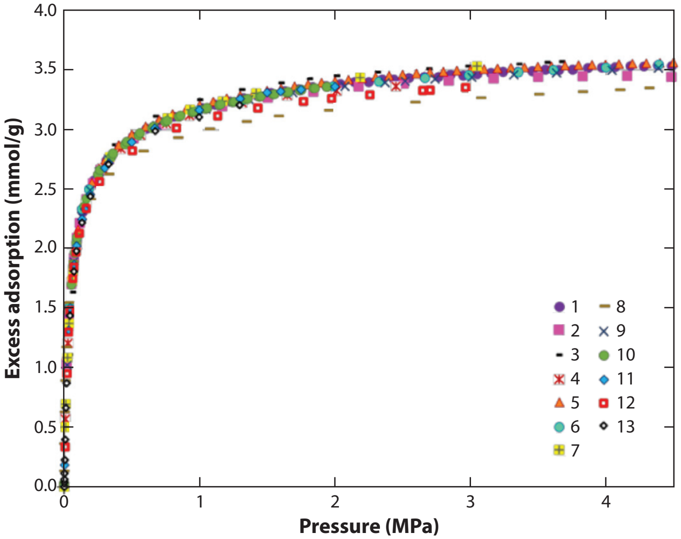

Figure 1

Thirteen independent measurements of excess adsorption of CO_2 in National Institute of Standards and Technology Reference Material RM 8852 (ammonium ZSM-5 zeolite) at 20°C made in 11 separate laboratories. Reproduced with permission from Reference 17.

involves many factors (16), the equilibrium adsorption isotherm for CO_2 is frequently used to judge whether a MOF (or another adsorbent) is useful or interesting. The topic Park and colleagues addressed was what can be definitively stated about the reproducibility of CO_2 adsorption isotherms in MOFs. The choice of this topic was based on the possibility of making firm statements about experimental reproducibility on a particular physical phenomenon, not because Park et al. believed that CO_2 adsorption in MOFs is in some way more (or less) reproducible than other physical phenomena that are widely studied in materials-oriented research.

A crucial aspect of Park et al.'s study is that the experimental measurement of CO_2 adsorption isotherms is relatively routine and can be performed with several widely available commercial instruments. This point was powerfully illustrated by a recent study led by the National Institute of Standards and Technology (NIST), in which 11 independent groups measured CO_2 adsorption at 20°C in a zeolite provided to each group by NIST (17). The results of the 13 experiments from this work, which were performed using a range of equipment, are shown in **Figure 1**. The good agreement between the experiments is evident. This study does not imply that making these measurements is trivial—care and skill are still needed to perform them accurately. Nonetheless, the results do indicate that at least in principle the reliable measurement of CO_2 adsorption in porous materials at room temperature is something that does not provide an enormous challenge to well-equipped researchers.

A second crucial aspect of the work of Park et al. is that a publically available database exists that has attempted to collect equilibrium adsorption isotherms (for all molecules, not just CO_2) from the open literature in a comprehensive way (18). Gathering these data required an enormous sustained effort by the NIST-led team that compiled it, and without it the work of Park et al. would have had a much more limited scope.

Park et al. analyzed the thousands of reported isotherms in the NIST Adsorption Database to collect all available measurements of CO_2 in MOFs. From these data, they identified all examples

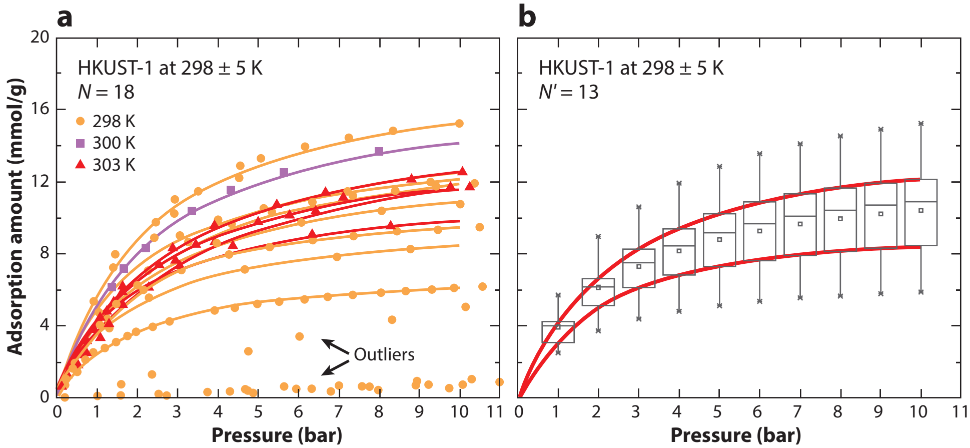

Figure 2

(*a*) Experimental data from 18 independent measurements of CO_2 adsorption in HKUST-1 at 298 ± 5 K, with temperatures indicated by color and symbol type. Outliers identified by the methods defined by Park et al. (11) are indicated. Solid curves show the fitted functions used in analysis of the data. (*b*) Box and whisker plot for 13 independent measurements of CO_2 adsorption in HKUST-1 at 298 ± 5 K obtained after rejecting outliers. Reproduced with permission from Park et al. (11), *Chem. Mater.* 29:10487–95; copyright 2017 American Chemical Society.

that had been reported multiple times. For a small number of materials, a relatively large number of replicates are available. The example with the most replicates, the well-known MOF HKUST-1, is shown in **Figure 2**. **Figure 2*a*** shows all 18 isotherms that are available for this material within a small temperature range. The contrast with **Figure 1** is clear. Park et al. defined some simple bias-free statistical methods to identify outliers among data such as **Figure 2*a***; 5 isotherms were identified as outliers using this approach. **Figure 2*b*** shows an example of a consensus isotherm that can be developed for CO_2 in HKUST-1 using the remaining 13 isotherms. It is important to note that the wide scatter in the data in **Figure 2** is not shocking to experts in synthesizing MOFs. HKUST-1 is known to be sensitive to moisture, so small variations in materials preparation and handling can lead to significant changes in the materials' properties (19).

Figures 1 and **2** show two strongly contrasting examples of repeated measurements of CO_2 adsorption in a nanoporous material, one of which is highly reproducible and the other of which is not. Which example is more representative? The work of Park et al. gives some insight into this question because they extended the approach outlined above to all reported CO_2 isotherms in MOFs. One of their key findings is that among all of the replicate isotherms that were analyzed, one in five (or more precisely, 21%) were classified as outliers. That is, one in five of the isotherms were statistically inconsistent with the other available replicates for the same material. This is a sobering observation, because if it is representative of the entire literature, it implies that selecting a single isotherm from the literature comes with a roughly one-in-five chance that the data differ very strongly from what would typically be considered the right answer. It would obviously be very interesting to establish whether this estimate is also a reasonable one for other areas of materials-oriented research.

The 20% occurrence of outliers discussed above would not be worrisome if there were access to enough replicate measurements that outliers could readily be detected. Unfortunately, this does not describe the reality of the literature associated with materials-oriented research. Although thousands of papers have been published with data on gas adsorption in MOFs, Park et al. found

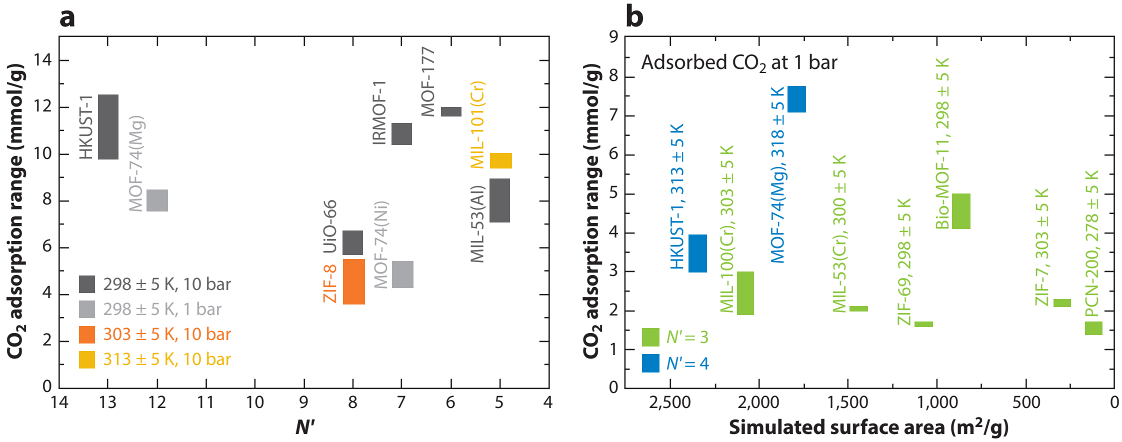

Figure 3

(*a*) Summary of interquartile range for CO_2 adsorption at 10 bar (1 bar for MOF-74) for all known metal-organic frameworks (MOFs) with more than four reported replicates after eliminating outliers. Numbers of independent measurements that exist after discarding outliers (N') were used for each material on the horizontal axis. (*b*) Summary of range for CO_2 adsorption at 1 bar for all known MOFs with $N' = 3$ or 4. Simulated surface areas were used for each material on the horizontal axis. Reproduced with permission from Park et al. (11), *Chem. Mater.* 29:10487–95; copyright 2017 American Chemical Society.

that only nine materials exist for which four or more replicate CO_2 isotherms are available after removing outliers. The state of knowledge for all of these materials is summarized in **Figure 3**. For the great majority of CO_2 adsorption isotherms in the open literature, no independent replicates are available.

We emphasize that the results of Park et al. focused on gas adsorption in MOFs because of the availability of a comprehensive data collection, not because the authors believed that experiments with these materials are inherently more (or less) reproducible than measurements in any other area of materials-oriented research. Our summary of their results should be viewed only as an initial case study on the reproducibility of research areas with the characteristics we listed in Section 1. A range of issues exist regarding what can and cannot be deduced from retrospective literature analysis of this kind, and we refer interested readers to the original work of Park et al. (11) for more details.

4. CASE STUDY 2: REPLICATED SYNTHESIS OF CRYSTALLINE NANOPOROUS MATERIALS

The first case study focused on a specific application of nanoporous materials, namely, adsorption of CO_2. This specific application may underestimate what is known about reproducibility of the underlying materials because subsequent work may follow the material synthesis originally reported but test this material for different applications. This suggests a more foundational question that can be asked about reproducibility of materials-oriented research: Once the synthesis of a new material is reported, how often is the synthesis independently replicated? If new materials in the literature are not made again (or more precisely, their synthesis is not reported again in the open literature), then it inevitably follows that nothing can be known about the reproducibility of these materials or their properties.

To probe a specific example of the question we posed above, we examined the literature describing the synthesis of MOF materials. Tens of thousands of papers have been published about MOFs and their properties, so it is difficult to argue that the community has not had ample opportunity to repeat the synthesis of materials. Although it would be fascinating to attempt a comprehensive analysis of this literature, doing so is far beyond the scope of our discussion here. Instead, we chose a small subset of materials and examined their history in the literature. As a result, our discussion in this section should be considered as anecdotal rather than as offering a definitive analysis.

Crystal structures of new crystalline materials such as MOFs are typically reported in the Cambridge Structural Database (CSD) (20) during or after publication. We selected 16 MOFs from the Computation-Ready, Experimental (CoRE) MOF database (21), which is a subset of the CSD containing nearly all disorder-free MOFs. Each of these 16 materials was initially reported between 2007 and 2013, and where possible we chose examples in which the original report described a single material. The latter choice greatly simplified analysis of later papers that cited the original results. At the time of writing, the 16 original papers had been cited between 8 and 168 times, with an average of approximately 50 citations. There may of course be legitimate reasons to not replicate studies of a particular material; for example, the performance of the material when it is first tested for a particular task might be poor. If a paper is cited many times, this is at least weak evidence that the material that is reported has enough value that the community is interested in it.

We examined each of the 768 papers that cited one of the original 16 synthesis reports to determine which materials had been synthesized again. We did not make judgments about the quality or outcome of repeat syntheses; if authors stated that they had made the same material as in an earlier paper, then we considered this a repeat synthesis. We considered chemical variants of a MOF sharing the same name, such as MIL-96 (Al) (22) and MIL-96 (Cr) (23), as reproduced syntheses. A citation was considered to have performed a modified synthesis if the organic linker was functionalized, if a substituted metal center fundamentally changed the primary property of interest (for example, the crystal structure), or if guest inclusions significantly changed the primary property of interest. Our discussion below does not include double counting; if a citing paper repeated an exact synthesis and then further modified it, we counted this as an example of an exact repeat synthesis but did not include the modified synthesis. We also examined the citing papers to see if they shared one or more authors with the original report. The results of this analysis are summarized in **Table 1** and **Figure 4**.

Of the 16 materials we surveyed, more than a quarter (5/16) have not been synthesized again by any authors, and more than half (9/16) have not been synthesized by a group of authors independent of the original authors. In other words, nothing can be concluded from the published literature about the reproducibility of the majority of the materials at even the minimal level of having the material made by two independent groups. Only three materials had their syntheses repeated exactly more than twice, and for only one of these materials were these replicates performed by researchers independent of the original authors. Interestingly, the number of reported replicates does not correlate strongly with the number of citations of the initial paper. For example, the material with the most replicates by original authors (material #3 in **Table 1**) was only the fifth most cited paper among the set we examined.

Perhaps the most striking observation from our results is that the number of modified syntheses that are reported is significantly larger than the number of exact syntheses. To give one example, material #1 has 2 exact syntheses reported in the literature to date, but modified syntheses have been reported 25 times. Recall that in our notation, a material was recorded as having a modified synthesis reported only if that report did not include data on synthesis of the original material. It seems likely that in many of the studies that developed modified syntheses, the original material was also produced during the course of the new work. This suggests that authors are consciously choosing to not include information about replicate syntheses when they report their work. If

Table 1 Analysis of citations from 16 metal-organic frameworks originally reported between 2007 and 2013[a]

Material label	Year	Citations	CSD code and reference	# Repeat syntheses (some author overlap)	# Repeat syntheses (new set of authors)	# Modified syntheses (some author overlap)	# Modified syntheses (new set of authors)
1	2007	77	IJOMOJ06 (24)	2	0	15	10
2	2007	8	XUBJZF02 (25)	1	1	0	0
3	2007	54	HOMZEP (26)	5	4	0	0
4	2007	15	DIXHIC (27)	0	0	0	0
5	2007	19	TICPOL (28)	1	1	0	0
6	2008	13	GOGSIF (29)	1	0	4	1
7	2009	37	GUKZAO (30)	0	0	0	0
8	2009	17	RUCGOM (31)	2	0	5	0
9	2010	86	XUNGUJ (32)	2	1	7	5
10	2010	27	QUQGEP (33)	1	0	0	0
11	2010	33	UKUBUY (34)	0	0	1	1
12	2010	27	MACHIJ (35)	1	1	0	0
13	2010	46	OSOYUR (36)	0	0	1	1
14	2010	104	MUVJIX (37)	2	1	10	7
15	2012	168	SAPBIW (38)	7	2	9	5
16	2013	37	RIDGIW (39)	0	0	0	0

[a]For each material the Cambridge Structural Database (CSD) code and a reference to the original paper are given. The terminology used for repeat and modified syntheses is defined in the text.

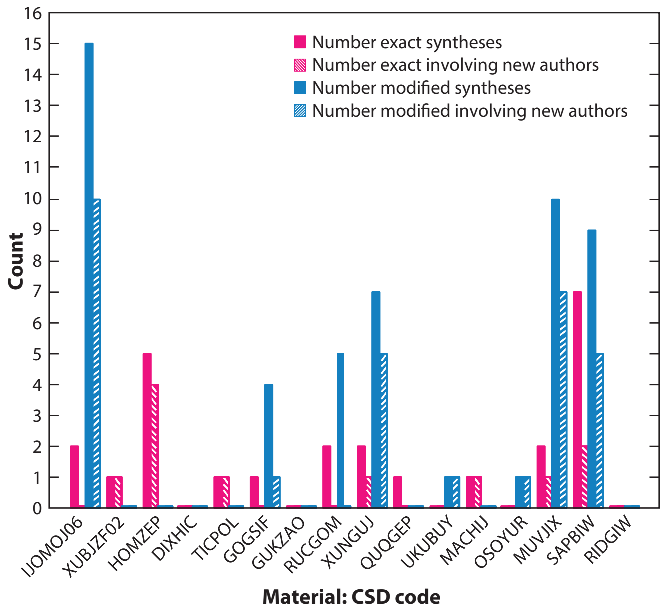

Figure 4

Materials synthesis reported in papers citing original reports of the 16 metal-organic frameworks described in the text. Abbreviation: CSD, Cambridge Structural Database.

this is correct, the research community is missing the opportunity to strengthen the reliability of published literature by reporting the replicate experiments.

The anecdotal data we have presented, although thought provoking, should be considered illustrative rather than definitive. We examined only a small number of examples chosen in a non-random way. We hypothesize that these results for synthesis of MOFs are indicative of what would be found by applying the same approach to other classes of materials, but we have not tested this hypothesis. Our analysis examines only whether materials have been synthesized repeatedly, and not the issue of whether the replicate synthesis gave consistent results. This latter issue is challenging to address quantitatively for MOF synthesis because the primary tool for assessing synthesis outcomes, powder X-ray diffraction, is in general a qualitative rather than quantitative tool.

5. DISCUSSION AND RECOMMENDATIONS

The preceding sections have been intended to prompt the reader to ponder the state of reproducibility in their own field of research. At an idealistic level, it is easy to conclude that the long-term impact and value to society of materials-oriented research would be improved if the reliability of information in the literature were higher, or even if it were simply easier to assess. At the same time, any effort to improve reproducibility comes with real costs in time and resources relative to current norms. In this section, we turn to the important question of what can be done to improve upon the current state of affairs. The suggestions we give are our personal opinions, not ideas that have developed as a strongly held consensus in the research community.

Concerns about reproducibility in science have led to a range of recommendations for improvements by individual investigators and at community levels by institutions, journals, and funding agencies (40–42). Although many of these recommendations have broad applicability, the differences in research culture between life sciences (e.g., biomedicine), behavioral sciences (e.g., psychology), and materials-oriented chemical engineering mean that recommendations developed in the former fields do not always translate well into the latter. In biomedicine and behavioral sciences, for example, an important recommendation is to increase the sample size of studies to strengthen the statistical power of information that can be derived. It would shock many scientists from these fields to learn that a typical measurement reported in a materials-based engineering paper has $N = 1$. That is, it is common to report data from a single sample or experiment. An appropriate response to this realization is to aim to quantify sources of uncertainty (by error bars or other methods) in key measurements. Stated this way, this suggestion seems uncontroversial in the extreme, and yet the senior authors of this article can easily find examples in our published work in which this effort was not explicitly reported. It would be a useful exercise to consider whether this is also the case for published work from your own research group.

A second recommendation that has been widely discussed in clinical and behavioral sciences is to preregister hypotheses to be tested and protocols to be followed before a study is started. This approach provides a powerful antidote to knowing or unknowing efforts to "p-hack" data or adjustments to data-handling methods on the fly to obtain interesting results. It is difficult, however, to translate this important concept into materials-oriented research for the following simple reason: Despite what our grant applications might say, much materials-oriented research is not truly hypothesis driven in the formal, statistically testable sense used in high-quality clinical research. A simplistic description of a typical set of materials-oriented experiments is that a set of known variables (e.g., identity of chemical functional groups) are varied and the influence of this variation on a group of physical properties (e.g., conductivity, gas adsorption) is measured. This description certainly implies an underlying hypothesis that the variables may or may not change the physical properties, but the experimenter's aim is almost always to quantify the properties, and

Table 2 The medal stand of replication in experiments with nonbiological materials

Medal	Description
Bronze	A single experimenter or team synthesizes a material more than once and measures the material's properties more than once
Silver	Two or more independent experimenters or teams measure the properties of a material using samples obtained from a single source (e.g., synthesis of a material by a single person)
Gold	Two or more experimenters or teams independently synthesize and measure properties of a material

not to test this hypothesis explicitly. This may seem like a semantic issue, but we contend that the concept of preregistering hypotheses and protocols would not be productive in materials-oriented research because of these accepted cultural norms. We hasten to add, however, that clear reporting of how data are gathered and processed is vital, as is the need to maintain high standards of evidence on claims that changing variable x causes important changes in physical property y.

Below, we give three broad recommendations specifically aimed at experimental research with nonbiological materials. The first two are aimed at individual researchers, whereas the third will require action by research communities. The two recommendations for individuals revolve around replicating experiments, so it is useful to briefly discuss what we mean by this term. A useful hierarchy of replicate experiments is listed in **Table 2** using the easy-to-understand structure of Olympic medals. The gold medal designation in **Table 2** is equivalent to the standard of repeated synthesis by independent authors discussed in Section 4. Recall that fewer than half of the widely cited materials we examined in Section 4 reached this standard. Many would argue that the situation we refer to as a bronze medal for replication is simply repeating experiments in one's own lab and should be a standard part of scientific norms and, thus, not worthy of a medal. However, there are many examples where this standard is not met in published papers, so it seems worthwhile to recognize it. In the recommendations below, we use replication to indicate the stronger forms of replication designated with silver and gold medals in **Table 2**. Despite the great value of replicates from these categories to the scientific community, their existence is rare enough that associating them with the honor of medals seems appropriate.

5.1. Recommendation 1: Perform Replication Experiments

Given the value of replicate experiments to directly assess reproducibility, this recommendation is obvious. It is useful to consider why this recommendation is not already widely followed. A source of resistance may be a scientific culture that values novelty and newness and views repeating past work as a suboptimal use of time. Although a high-minded appeal to scientific principles is tempting to counter this view, we instead make three observations that provide positive reinforcement to researchers making use of their scarce resources:

- New students need to be trained, and asking them to repeat previous experiments that have well-defined methods and analysis is an excellent approach to training. The benefits of teaching students what it means to perform high-quality experiments, critically read the literature, and produce publication-quality results early in their career are enormous. A simple way to bolster this approach within your own institution is to ask about replication routine in PhD proposals, PhD defenses, seminars, job interviews, and similar venues.
- Many researchers, including ourselves, are at least partially funded by sources that use the total number of papers as one metric of success. There are of course powerful arguments to make against using simplistic metrics such as the number of papers to make sophisticated

scientific judgments. Researchers who seek to have long-term impact should seek to write substantial papers that truly change a field rather than thinly dividing work into minimal publishable units. Nevertheless, the reward structure of our scientific culture is often geared to reward numerical productivity, and publishing replication efforts that bring value to the scientific community can contribute to this kind of productivity.

- Journals from high-quality publishers now exist that explicitly aim to publish scientifically sound work without concern for the study's "impact" (43). This eliminates the legitimate concern that if a replication study is submitted to a high-profile journal, it will be rejected because it is perceived to have limited impact.

The second and third observations above lead directly to our second recommendation.

5.2. Recommendation 2: Publish the Results of Replication Experiments

A replicated set of experiments is valuable to the scientific community only if the results are published. To encourage researchers, including ourselves, to publish results of this kind, we present three scenarios in which this would be appropriate.

- Include replication experiments in a manuscript that deals primarily with new materials or effects. Replicate experiments of this kind are often performed during the initial stages of work on new materials, but they are frequently not discussed in print. We provided anecdotal evidence for this state of affairs in Section 4. It would be valuable to learn to think of these results as something that should be published. The widespread availability of supporting information means that it cannot be argued that including these results would make a manuscript too long.
- Publish manuscripts that are solely devoted to replication experiments. As noted above, high-quality venues now exist in which publications of this type would be considered without prejudice. If replicating experiments is made a standard part of student training, then publishing manuscripts of this kind offers a natural way to extend this training to preparation and completion of a manuscript.
- Publish, with care, negative results. Replication efforts that contradict the original study may contain more valuable information than positive outcomes and should be viewed as a key pillar of science's self-correcting state. Nonetheless, there are special challenges associated with publishing outcomes of this kind. At the very least, it may be likely that one of the authors of the original study will be asked to review the manuscript. We have several suggestions for navigating this situation. First, conducting the relevant experiments with additional characterization or supporting calculations would be appropriate. Second, reporting replication experiments for a set of materials from the literature where (presumably) some of the outcomes are positive may be a way to include a negative result. This is certainly less confrontational than the approach of writing a paper with a title like "Investigator A's Experiments with Material B Were Wrong." Finally, scrupulously stick to facts in reporting your results, and do not ascribe motives to the previous study.

5.3. Recommendation 3: Establish Material Standards and Enforce Their Use as a Required Element of Publishing New Results

A common approach in materials-oriented research is to compare the measurements of materials from different studies. However, these comparisons are typically made without the benefit of replicate experiments that can establish different studies as appropriate apples-to-apples comparisons. There are many situations in which details of experimental procedures (for example, the presence of trace moisture or other species during sample preparation or handling) are not always reported. In these instances, drawing meaningful conclusions from comparisons of literature data can be difficult. A potential solution to this problem is for technical communities to establish readily available material standards for particular measurements and to create a culture in which the use of these standards is expected for data from new materials to be published. Having a standard of this kind requires that specific materials be either readily available or readily synthesized based on independent experimental results from multiple groups. We gave a specific example of this approach in Section 3, where a NIST-led effort performed round-robin measurements of CO_2 adsorption at room temperature by multiple independent experimental groups using a well-defined material that is now readily available to researchers worldwide (17). Within the narrowly defined community interested in CO_2 adsorption in porous materials, we feel that including data showing measurements consistent with the consensus results for this material should become a de facto standard for having a new manuscript published. Although these kinds of standards can in principle be enforced by journals, it is likely to be more effective for individual reviewers to take this action, as they are better equipped to understand the norms within narrowly defined subtopics of research.

A downside of this third recommendation is that it requires concerted action and additional resources. The example we gave, after all, was associated with extensive effort made by researchers at a government-funded organization entirely focused on technological standards (NIST), not to mention time and effort by many additional research labs. However, once a standard of this kind is established, it is almost guaranteed to be extensively cited. This provides a significant positive incentive for self-organized efforts to establish new standards. This is an area in which early- or mid-career researchers can be particularly effective. We recommend this as a topic for conversation at the next conference coffee break or lunch you attend as an alternative to lamenting the indignities of travel or your institution's accounting department.

CONCLUSION

In this article, we have focused on the issue of reproducibility as it relates to research on experiments with nonbiological materials. Many of the same issues also arise in research on computational modeling of these materials. In principle, reproducibility should be easier to achieve in computational research, where direct access to the codes, input files, and so forth used in an initial study is possible. Some subfields now make working versions of codes available to readers (44), but this approach is unlikely to work in fields that use commercial or proprietary codes. Specific suggestions for enhancing the reproducibility of computational research, as well as ideas for making computational research more relevant to a broad community, have been discussed elsewhere (45).

It is easy when thinking about scientific reproducibility to focus on the substandard approaches of other people. It is far more important, however, for each of us to think clearly about our own work. We hope that reading this article has prompted you to think in new ways about the short- and long-term goals of your own research. Whether you are a new graduate student or a world-leading senior researcher, there are constructive steps you can take to increase the reproducibility (and therefore, impact) of your own work and that of those around you. We encourage you to consider the recommendations above as a starting point for discussion with your collaborators and co-workers and to actively explore how they can be adapted or expanded to improve the overall quality of work in your specific domain.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work is supported by UNCAGE-ME, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under Award #DE-SC0012577.

LITERATURE CITED

1. Harris R. 2017. *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions*. New York: Basic Books
2. Ioannidis JP. 2005. Why most published research findings are false. *PLOS Med*. 2:e124
3. 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
4. Klein RA, Ratliff KA, Vianello M, Adams R, Bahník Š, et al. 2014. Investigating variation in replicability: a "many labs" replication project. *Soc. Psychol.* 45:142–52
5. 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
6. Begley CG, Ellis LM. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* 483:531
7. Fanelli D. 2018. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *PNAS* 115:2628–31
8. Natl. Renew. Energy Lab. 2018. *Photovoltaic Research: Device Performance*. Golden, CO: Natl. Renew. Energy Lab. **https://www.nrel.gov/pv/device-performance.html**
9. Walton KS, Sholl DS. 2017. Research challenges in avoiding "showstoppers" in developing materials for large-scale energy applications. *Joule* 1:208–11
10. Mahaffey J. 2017. *Atomic Adventures: Secret Islands, Forgotten N-Rays, and Isotopic Murder: A Journey into the Wild World of Nuclear Science*. New York: Pegasus Books
11. Park J, Howe JD, Sholl DS. 2017. How reproducible are isotherm measurements in metal-organic frameworks? *Chem. Mater.* 29:10487–95
12. Eddaoudi M, Kim J, Rosi N, Vodak D, Wachter J, et al. 2002. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* 295:469–72
13. 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
14. Cohen SM. 2012. Postsynthetic methods for the functionalization of metal-organic frameworks. *Chem. Rev.* 112:970–1000
15. 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
16. Maring BJ, Webley PA. 2013. A new simplified pressure/vacuum swing adsorption model for rapid adsorbent screening for CO_2 capture applications. *Int. J. Greenh. Gas Control* 15:16–31
17. Nguyen HGT, Espinal L, van Zee RD, Thommes M, Toman B, et al. 2018. A reference high-pressure CO_2 adsorption isotherm for ammonium ZSM-5 zeolite: results of an interlaboratory study. *Adsorption* 24:531–39
18. Siderius DW, Shen VK, Johnson RD III, van Zee RD. 2015. *NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials*. Gaithersburg, MD: Natl. Inst. Stand. Technol.
19. Burtch NC, Jasuja H, Walton KS. 2014. Water stability and adsorption in metal-organic frameworks. *Chem. Rev.* 114:10575–612
20. Groom CR, Allen FH. 2014. The Cambridge Structural Database in retrospect and prospect. *Angew. Chem. Int. Ed.* 53:662–71
21. Chung YG, Camp J, Haranczyk M, Sikora BJ, Bury W, et al. 2014. Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.* 26:6185–92
22. Volkringer C, Popov D, Loiseau T, Férey G, Burghammer M, et al. 2009. Synthesis, single-crystal X-ray microdiffraction, and NMR characterizations of the giant pore metal-organic framework aluminum trimesate MIL-100. *Chem. Mater.* 21:5695–97

23. Long P, Wu H, Zhao Q, Wang Y, Dong J, Li J. 2011. Solvent effect on the synthesis of MIL-96(Cr) and MIL-100(Cr). *Microporous Mesoporous Mater*. 142:489–93
24. Zhang B, Wang ZM, Kurmoo M, Gao S, Inoue K, Kobayashi H. 2007. Guest-induced chirality in the ferrimagnetic nanoporous diamond framework $Mn_3(HCOO)_6$. *Adv. Funct. Mater.* 17:577–84
25. Zhang C-Z, Mao H-Y, Wang Y-L, Zhang H-Y, Tao J-C. 2007. Syntheses of two new hybrid metal-organic polymers using flexible aliphatic dicarboxylates and pyrazine: crystal structures and magnetic studies. *J. Phys. Chem. Solids* 68:236–42
26. Volkringer C, Loiseau T, Férey G, Morais CM, Taulelle F, et al. 2007. Synthesis, crystal structure and ^{71}Ga solid state NMR of a MOF-type gallium trimesate (MIL-96) with μ3-oxo bridged trinuclear units and a hexagonal 18-ring network. *Microporous Mesoporous Mater*. 105:111–17
27. Chang W-M, Cheng M-Y, Liao Y-C, Chang M-C, Wang S-L. 2007. Template effect of chain-type polyamines on pore augmentation: five open-framework zinc phosphates with 16-ring channels. *Chem. Mater.* 19:6114–19
28. Schull TL, Henley L, Deschamps JR, Butcher RJ, Maher DP, et al. 2007. Organometallic supramolecular mixed-valence cobalt(I)/cobalt(II) aquo complexes stabilized with the water-soluble phosphine ligand *p*-TppTp (*p*-triphenylphosphine triphosphonic acid). *Organometallics* 26:2272–76
29. Kiskin MA, Aleksandrov GG, Bogomyakov AS, Novotortsev VM, Eremenko IL. 2008. Coordination polymers of cobalt(II) with pyrimidine and pyrazine: syntheses, structures and magnetic properties. *Inorg. Chem. Commun.* 11:1015–18
30. He J, Yang C, Xu Z, Zeller M, Hunter AD, Lin J. 2009. Building thiol and metal-thiolate functions into coordination nets: clues from a simple molecule. *J. Solid State Chem.* 182:1821–26
31. Konno T, Yoshinari N, Taguchi M, Igashira-Kamiyama A. 2009. Drastic change in dimensional structures of D-penicillaminato $(Au^{I}_2Pt^{I}_2Zn^{II})_n$ coordination polymers by moderate change in solution pH. *Chem. Lett.* 38:526–27
32. Abrahams BF, Grannas MJ, Hudson TA, Robson R. 2010. A simple lithium(I) salt with a microporous structure and its gas sorption properties. *Angew. Chem. Int. Ed.* 49:1087–89
33. Hu B-W, Zhao J-P, Tao J, Sun X-J, Yang Q, et al. 2010. A new azido-nickel compound with three-dimensional Kagomé topology. *Cryst. Growth Design* 10:2829–31
34. Zhang J, Xue Y-S, Liang L-L, Ren S-B, Li Y-Z, et al. 2010. Porous coordination polymers of transition metal sulfides with PtS topology built on a semirigid tetrahedral linker. *Inorg. Chem.* 49:7685–91
35. Yue Q, Yan L, Zhang J-Y, Gao E-Q. 2010. Novel functionalized metal-organic framework based on unique hexagonal prismatic clusters. *Inorg. Chem.* 49:8647–49
36. Zou R, Zhong R, Han S, Xu H, Burrell AK, et al. 2010. A porous metal-organic replica of α-PbO_2 for capture of nerve agent surrogate. *J. Am. Chem. Soc.* 132:17996–99
37. Blake AJ, Champness NR, Easun TL, Allan DR, Nowell H, et al. 2010. Photoreactivity examined through incorporation in metal-organic frameworks. *Nat. Chem.* 2:688
38. An J, Farha OK, Hupp JT, Pohl E, Yeh JI, Rosi NL. 2012. Metal-adeninate vertices for the construction of an exceptionally porous metal-organic framework. *Nat. Commun.* 3:604
39. Zhang HX, Fu HR, Li HY, Zhang J, Bu X. 2013. Porous ctn-type boron imidazolate framework for gas storage and separation. *Chemistry* 19:11527–30
40. Natl. Acad. Sci. Eng. Med. 2016. *Statistical Challenges in Assessing and Fostering the Reproducibility of Scientific Results: Summary of a Workshop*. Washington, DC: Natl. Acad. Press
41. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, et al. 2014. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 383:166–75
42. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, et al. 2016. Increasing value and reducing waste in biomedical research: Who's listening? *Lancet* 387:1573–86
43. Bohne C, Liz-Marzán LM, Ganesh KN, Zhang D. 2016. Chemistry, from alpha to omega, open to all. *ACS Omega* 1:1
44. Torfi A, Iranmanesh SM, Nasrabadi N, Dawson J. 2017. 3D convolutional neural networks for cross audio-visual matching recognition. *IEEE Access* 5:22081–91
45. Zuilhof H, Yu S-H, Sholl DS. 2018. Writing theory and modeling papers for *Langmuir*: the good, the bad, and the ugly. *Langmuir* 34:1817–18