YALE JOURNAL OF BIOLOGY AND MEDICINE 90 (2017), pp.147-151.

YJBM

Is Open Science the Future of Drug Development?

Daniel L. Shaw*

Yale School of Medicine, Yale University, New Haven, CT

Traditional drug development models are widely perceived as opaque and inefficient, with the cost of research and development continuing to rise even as production of new drugs stays constant. Searching for strategies to improve the drug discovery process, the biomedical research field has begun to embrace open strategies. The resulting changes are starting to reshape the industry. Open science—an umbrella term for diverse strategies that seek external input and public engagement—has become an essential tool with researchers, who are increasingly turning to collaboration, crowdsourcing, data sharing, and open sourcing to tackle some of the most pressing problems in medicine. Notable examples of such open drug development include initiatives formed around malaria and tropical disease. Open practices have found their way into the drug discovery process, from target identification and compound screening to clinical trials. This perspective argues that while open science poses some risks—which include the management of collaboration and the protection of proprietary data—these strategies are, in many cases, the more efficient and ethical way to conduct biomedical research.

In 1991, a 21-year-old Finnish student named Linus Torvalds posted on an early Internet message board, describing a "free" operating system that he was developing as "a hobby" [1]. The system, championed by Torvalds but developed virtually with additions and changes by innumerable individuals to its openly available source code, has since become Linux. A dominant platform in many computational environments, Linux represents a paradigm shift in development methods and is perhaps the single most recognizable product of the open source movement. The success of this open source software serves as one example of the ways in which decentralized research and development can be applied. Such open research models are now finding their way into diverse industries including pharmaceutical sciences [2].

In the pharmaceutical industry, practical limitations, cultural norms, and intellectual property concerns have resulted in a development process traditionally conducted in a carefully guarded proprietary environment [3-6]. Faced with rising costs and a reduced output of novel drugs, however, developers are choosing to selectively leverage open science models to improve the process of biomedical innovation. Such changes have been met with a mixture of excitement and skepticism, but are nonetheless occurring in many settings and stages of drug discovery—from large corporations to individual academic investigators, and from target identification to clinical trials. A few examples: Pharmaceutical giant Eli

*To whom all correspondence should be addressed: Daniel L. Shaw, BA, Yale School of Medicine, Yale University, 333 Cedar Street, New Haven, CT 06510, daniel.shaw@yale.edu.

†Abbreviations: R&D, Research and Development; NME, New Molecular Entity; NBE, New Biologic Entity; FDA, Food and Drug Administration.

Keywords: Drug Discovery, Drug Industry, Pharmaceutical Research, Research Design, Open Science, Crowdsourcing, Research Ethics, Reproducibility of Results

Author contributions: This article and the opinions expressed are the sole work of Daniel L. Shaw.

Lilly has spun out an online crowdsourcing platform [7], a public-private partnership working on antimalarials is using open source principles in compound screening [8], academic chemical biologists are sharing valuable early stage drug-molecules [9,10], and outcomes researchers are deploying clinical trial data-sharing infrastructures at Yale's Open Data Access project [11].

To be sure, biomedical scientists have a long history of embracing certain aspects of open science—a term that for the purposes of this article includes the full range of scientific work that seeks external input and public engagement. Peer-reviewed scientific journals and project-oriented collaborations such as the Human Genome Project are two examples. At the same time, risks embodied in open enterprise include both increased competition and the possible loss of proprietary discoveries [12-14]. While there may be significant synergies in effectively leveraging open research, there is no denying a potential clash between the often high-minded goals of open science and the highly-competitive culture of drug development. As such, the way that the pharmaceutical industry will ultimately employ technology to create a new era of open science is still uncertain. The questions generated by the rise of open drug development, however, are already important.

DO THE CURRENT DEVELOPMENT MODELS NEED TO BE UPDATED?

The drug development process is built on advances in basic scientific knowledge, with implicit interest in maximizing both innovation and efficiency [15]. The research and development (R&D†) enterprise is traditionally performed in-house at fully integrated large corporations, with input from academics and regulatory agencies as necessary [4,16,17]. While innovation is at the heart of the drug development process, such innovation can be diffuse, making it hard to quantify or define. For one, the drug discovery process is lengthy. It may take decades for basic science discovery to be translated into life-saving therapy [18]. It is also incredibly difficult to predict which drug or drug class will eventually make it to market [19]. Even high-profile research can prove unreliable, with some studies showing that many landmark papers are irreproducible [20]. Given these complications, the output of discovery efforts are most easily quantifiable as novel drug therapies—new molecular entities (NMEs) or new biological entities (NBEs)—approved by the US Food and Drug Administration (FDA). However, this approach captures neither the novelty nor the impact on human health of new drugs.

Examination of the efficiency of the process is more straightforward. It can be determined by calculating inputs such as funding, personnel, and scientific knowledge divided by outputs such as NMEs. In 2016, the National Institutes of Health spent $30 billion and industry spent at least $50 billion on R&D [21,22]. Much of it went towards innovation as represented by the development of NMEs and NBEs. Using this metric, simple models show that the absolute cost in monetary terms has increased exponentially over the last decades even as the total number of new drugs has remained relatively constant [4,23,24]. A recent nuanced analysis explored the numerous inputs required to produce NMEs, including funding, personnel, and scientific knowledge (papers/patents) [15]. The results: Since 1965, the number of scientific publications has increased 527 percent annually and number of authors has increased 807 percent annually. At the same time, the decade from 2004 to 2014 saw a marked decrease in drug approval, while research efficiency fell to the lowest level in modern drug development history.

These statistics undermine a commonly accepted premise that modern drug discovery is scalable, linear, and inevitable. Rather, the process may be far more complex and recursive than current models suggest. Drug development may not proceed in a straight line from target identification to drug screening to optimization to clinical trials. Reflecting that reality, emerging models are being designed to provide greater flexibility, improved collaboration, and increased pre-clinical research that may reduce needless human study [3,16]. Open science, in particular, has been touted as offering more opportunistic models [3,5,12].

HOW OPEN IS OPEN?

Efforts to incorporate open science into R&D may include goal-directed collaboration, data sharing, crowd-sourcing, and open source initiatives [12]. Such research transparency may open new avenues for discovery that create value. Each of these methods is now part of the drug development ecosystem, yet not all are appropriate for every research endeavor.

There are risks: collaborators or competitors can scoop publications, file patents that impede commercialization, and create management problems when personalities clash [3,5,12-14]. Moreover, some researchers may not benefit from opening the doors prematurely to outside input. Preliminary studies of a new topic, early stage preclinical development, and Phase 1 clinical trials may be inopportune times to engage outside partners. When seeking a change in direction or rapid development along a chosen path, however, there may be real benefit in strategically opening the research.

Goal-Directed Collaboration: Collaboration is a standard and well-established model of opening scientific research. Teams offering varying expertise may find groups to share in the discovery process. Academic labs,

for example, may team up with for-profit initiatives, and larger consortiums can establish networks for collaboration. Public-private partnerships and development deals are a common mechanism for bringing to market novel medications for under-studied disease processes [3,5]. Successful collaborations typically require strong leadership and may be focused around a particular goal, whether it be a disease such as malaria or a discovery process such as the human genome project.

Open data: Open data involves sharing the data resultant from scientific inquiry with the broader research community. Examples include the sharing of genome sequences, protein crystal structures, compound screening results, and clinical trial data. The scientific community is then able to engage with and build off this data. In an open data model, originators maintain complete control over the production of the data, while others have access but may not alter the original data sets. Open data is already the standard practice in numerous fields. Computational biology and structural biology commonly deposit their data [25] or code in public repositories. In other fields, data sharing is still in the early stages. Sharing clinical trial data is another promising area, and many clinical trials are now required to deposit key data on ClinicalTrials.gov [26]. However, the sharing of patient information is fraught with both legal and ethical concerns.

Crowdsourcing: With the power of massive computing and networking, it is now possible for informal collaboration to occur around targeted discovery efforts. There are a number of crowd-based games that involve individuals performing isolated pieces of research that contribute to computational biology modelling of protein-folding or genomics [27]. Additionally, large drug manufacturers have started crowd-sourcing networks to get outside input on challenging science. Website InnoCentive was established by drug giant Eli Lilly and then spun-out as a company for crowdsourcing science. It poses scientific challenges to the general public, offering prize money as an incentive for people to submit innovative solutions. Another example is the website Scientist.com, which seeks to establish an online marketplace for scientific goods and services [7].

Open Source: Open source research initiatives rely on a research structure in which both inputs and outputs of the research endeavor remain available to the public. The definition of the open source model continues to evolve but currently includes 1) creation of a good, service, or product; 2) open access to consume and contribute; 3) centrality of interaction; 4) purposeful but loosely affiliated work [2]. In theory, this means that each step of the research process—from target identification to clinical trial—would remain in the public domain. Work is protected through licensing—such as creative commons—that appropriately articulates the expectations for contribution and use of resultant research [13]. Open source does not require thousands of individuals to contribute and may be successful with even a handful of committed participants.

WHEN TO OPEN?

In basic research, the concept of open science has been well explored. From open access journals such as PLoS [28] to data repositories, open science is clearly part of the evolving landscape. There continue to be challenges to incentivizing the open dissemination of knowledge and providing credit for work performed. Drug development is somewhat different in that the end-product of discovery may be more fungible than basic science. Moreover, different stages of discovery lend themselves to various types of collaborative effort, which evolve as trust and efficiency dictate. Given the long and complex road to drug discovery, there are innumerable options for how to open the process. They include:

Target identification and Screening: A critical part of the drug discovery process is the identification, often through screening assays, of both a biologically relevant target and a specifically-binding drug-like molecule. Traditionally, the outcomes of these screens are closely guarded trade secrets [3]. However, in certain cases, this knowledge is now being opened to the public. The most prominent examples involve partnerships organized around developing treatment for neglected disease. Malaria—a disease disproportionately affecting underdeveloped nations with little potential for profitable drug manufacturing—has become a nidus of early stage data sharing. For example, Medicines for Malaria Venture (MMV), which started as a private-public partnership, evolved into a product development partnership. Out of these efforts emerged the publicly-available "Malaria Box," a set of 400 chemotypes identified from a phenotypic screen against *Plasmodium falciparum* [8]. Tuberculosis has also been the target of open source development, with over a hundred new compound hits recently identified and shared with the public [29].

Tool Compounds and Pre-clinical Study: At the same time, compound-sharing has emerged as a uniquely powerful paradigm of open drug discovery. Collaborations around pharmacologically active compounds have long served as a powerful accelerator of drug development, from penicillin [30] to cholesterol-lowering statins [31]. Discovery chemistry, however, poses a unique challenge to would-be innovators since knowledge of the source code (chemical structure) does not guarantee replication of compound. Efforts are now being made to establish systematized access to these compounds, with some individual labs leading the charge by widely sharing novel biologically active "chemical probes" [9,10,32].

One factor slowing the spread of compound sharing is the myth that providing molecules to others somehow invalidates patents or reduces the value of discovery [17]. In fact, encouraging research on a promising chemical probe may increase the value of that discovery. Consider a biologically-validated compound that is shared with academic laboratories so that they can do meaningful research. This may result not only in basic science knowledge but also an improved understanding of the pharmacology necessary to bring that drug class to market. Compound-sharing also allows companies to leverage discoveries that they have been unable to develop further. Indeed, top drug companies, such as Procter & Gamble, are known to actively manage only about 10 percent of their patents [14]. The remaining 90 percent remain unused, despite potential value to both the patent holder and the general public.

Clinical Trials: In the arena of clinical study, open science takes a different form. Here policy and ethical considerations are already changing the way in which clinical study data is managed. The NIH and leading journals have argued that since the public contributes to human-subject research by participating in clinical studies, the public is therefore ethically entitled to the data produced [33-35]. Already, ClinicalTrials.gov serves as a repository for all existing clinical trials that fall under certain rules, which were recently expanded. This may reduce unnecessary duplication of data, and increase patient access. The US Department of Health and Human Services recently took steps to increase the transparency of these studies [36].

Additionally, many large trials are now releasing their data to researchers for secondary analysis. Programs such as ClinicalStudyDataRequest.com and Yale Open Data Access [11] have created portals and review systems for outside groups to access raw data. This allows valuable patient data to be more fully leveraged through fact-checking, secondary analysis, and thorough literature review. One can even imagine a more participatory and open form of clinical trial in which individual patients use their smart devices to engage with the study designers [6]. An economic analysis by McKinsey & Company shows some $500 billion in value could be unleased through clinical data liquidity [37]. Taken in aggregate, such technological improvement could be used to facilitate both transparency and discovery.

TOWARDS AN OPEN FUTURE?

Enabled by technological innovation, open science is now part of the drug development landscape. Yet not every endeavor is appropriate for an open approach. As such, biomedical scientists need to pay attention to both the benefits and risks of open models. Collaboration, crowd-sourcing, data sharing, and open-sourcing may not have a role in a given research endeavor today, but they will undoubtedly play a role in the future of drug development. Perhaps the biggest risk posed by open science is ignoring these rapidly changing development tools.

Acknowledgements: The work and mentorship of Dr. James Bradner and Dr. Joseph Ross introduced me to the possibilities inherent in the open science model. I wish to acknowledge their impressive contributions to open scientific research in the realms of drug discovery and clinical trials, respectively.

REFERENCES

1. Moon J. Sproull L. Essence of distributed work. Online Communication and Collaboration: A Reader, 2010:125.
2. Levine SS, Prietula MJ. Open collaboration for innovation: principles and performance. Organization Science, 2013.
3. Paul SM, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nature reviews Drug discovery. 2010;9(3):203-214.
4. Munos B. Lessons from 60 years of pharmaceutical innovation. Nature Reviews Drug Discovery. 2009;8(12):959-968.
5. Munos B. Can open-source R&D reinvigorate drug research? Nature Reviews Drug Discovery. 2006; 5(9):723-729.
6. Munos B. Biomedical Innovation: Lessons From the Past and Perspectives for the Future. Clinical Pharmacology & Therapeutics. 2016.
7. Howe J. The rise of crowdsourcing. Wired magazine. 2006; 14(6):1-4.
8. Spangenberg T, et al. The open access malaria box: a drug discovery catalyst for neglected diseases. PloS one. 2013;8(6):e62906.
9. Frye SV. The art of the chemical probe. Nature chemical biology. 2010;6(3):159-161.
10. Scott AR. Chemical probes: A shared toolbox. Nature. 2016;533(7602):S60-S61.
11. Krumholz HM, et al. A historic moment for open science: the Yale University Open Data Access Project and Medtronic. Annals of internal medicine. 2013;158(12):910-911.
12. Maurer SM. Open source drug discovery: finding a niche (or maybe several). UMKC L. Rev. 2007;76:405.
13. Robertson MN, et al. Open source drug discovery–a limited tutorial. Parasitology. 2014;141(01):148-157.
14. Alexy O, Criscuolo P, and Salter A. Does IP strategy have to cripple open innovation? MIT Sloan Management Review. 2009;51(1):71.
15. Bowen A and Casadevall A. Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. Proceedings of the National Academy of Sciences. 2015;112(36):11335-11340.
16. Kaitin, KI. Deconstructing the drug development process: the new face of innovation. Clinical pharmacology and therapeutics. 2010;87(3):356.
17. Hunter J and Stephens S. Is open innovation the way forward for big pharma? Nature Reviews Drug Discovery. 2010;9(2):87-88.
18. Contopoulos-Ioannidis DG, et al., Life cycle of trans-

lational research for medical interventions. Science. 2008;321(5894):1298-1299.
19. Contopoulos-Ioannidis DG, Ntzani EE, and Ioannidis JP. Translation of highly promising basic science research into clinical applications. The American journal of medicine. 2003;114(6):477-484.
20. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2(8):e124.
21. U.S. Office of Budget. HHS FY2016 Budget in Brief. 2016 [cited 2016 12/28/2016]; Available from: https://www.hhs.gov/about/budget/budget-in-brief/nih/index.html#budget.
22. Pharmaceutical Research and Manufacturers of America. 2016 biopharmaceutical research industry profile. 2016; Available from: http://phrma-docs.phrma.org/sites/default/files/pdf/biopharmaceutical-industry-profile.pdf.
23. DiMasi JA, Grabowski HG, and Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. Journal of health economics. 2016;47:20-33.
24. Morgan S, et al. The cost of drug development: a systematic review. Health Policy. 2011;100(1):4-17.
25. Berman HM, et al. The protein data bank. Nucleic acids research. 2000;28(1):235-242.
26. Zarin DA, et al. Trial reporting in ClinicalTrials. gov—the final rule. New England Journal of Medicine. 2016;375(20):1998-2004.
27. Ranard BL, et al. Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. Journal of general internal medicine. 2014;29(1):187-203.
28. Brower V. Public library of science shifts gears: As scientific publishing boycott deadline approached, advocates of free scientific publishing announce that they will create their own online, free-access archive. 2001.
29. Ballell L, et al., Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. ChemMedChem. 2013;8(2):313-321.
30. Ligon BL. Penicillin: its discovery and early development. in Seminars in pediatric infectious diseases. 2004. Elsevier.
31. Endo A. A historical perspective on the discovery of statins. Proceedings of the Japan Academy, Series B. 2010;86(5):484-493.
32. Shaw DL, et al. Abstract LB-2: Open innovation drug discovery: A study in JQ1. Cancer Research. 2013;73(8 Supplement):LB-2-LB-2.
33. De Angelis C, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. New England Journal of Medicine. 2004;351(12):1250-1251.
34. Rodwin MA and Abramson JD. Clinical trial data as a public good. JAMA. 2012;308(9):871-872.
35. Hudson KL and Collins FS. Sharing and reporting the results of clinical trials. JAMA. 2015;313(4):355-356.
36. Shaw D and Ross J. US Federal Government Efforts to Improve Clinical Trial Transparency with Expanded Trial Registries and Open Data Sharing. AMA journal of ethics. 2015;17(12):1152.
37. Kayyali B, Knott D, and Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey & Company, 2013:1-13.