Innovations

# Structural Approach to Assessing the Innovativeness of New Drugs Finds Accelerating Rate of Innovation

Todd J. Wills* and Alan H. Lipkus

Cite This: *ACS Med. Chem. Lett.* 2020, 11, 2114−2119

Read Online

ACCESS | ⬛ Metrics & More | 📰 Article Recommendations | ⓢ Supporting Information

**ABSTRACT:** Measuring innovation in the pharmaceutical industry is challenging. Counts of new molecular entities (NMEs) approved by the Food and Drug Administration (FDA) are commonly used, but this measure only gauges quantity not innovativeness. A new indicator of innovation for small molecule and peptide drugs based on structural novelty is proposed and used to analyze recent trends in pharmaceutical innovation. We show pharmaceutical innovation has significantly increased over the last several decades despite recent concerns over an innovation crisis and find Pioneers (a NME whose shape and scaffold were not used in any previously FDA-approved drugs) are significantly more likely to be the source of promising new therapies. Analysis of the underlying source of structural innovation indicates that scaffolds first reported in the CAS REGISTRY five or less years prior to their Investigational New Drug application (IND) or on scaffolds populated with 50 or less other compounds at the time of IND tend to be the main source of Pioneers. Our analysis also shows a widening structural innovation gap between large pharmaceutical companies (Big Pharma) and the rest of the ecosystem even though the number of Big Pharma originated Pioneers has increased.

**KEYWORDS:** Drug discovery, Food and Drug Administration (FDA) drug approvals, innovation, new molecular entities (NMEs), small molecules

Structural Innovation Matrix
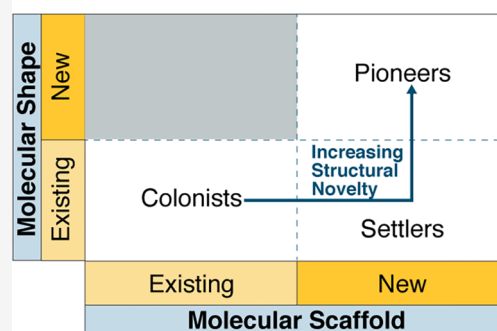
Although there has been little consensus as to what constitutes drug innovation, an innovation crisis has been a popular topic in the pharmaceutical industry over the past decade.[1,2] The total number of new molecular entities (NMEs) approved by the Food and Drug Administration (FDA) each year is a common benchmark used to measure the pace of innovation in pharmaceuticals.[3,4] However, a count based measure of innovation is focused on output and is not necessarily indicative of the innovativeness of the NMEs as not all NMEs are equally innovative.

Various attempts at assessing innovativeness have been made using such noncount based criteria as new mechanisms of action (first-in-class), therapeutic need (orphan drug), or improvement over the existing standard of care (breakthrough therapy). Depending on the definition utilized, various studies have found a positive or negative trend in drug innovation. Studies focused on first-in-class and orphan drugs have reported increasing innovation as first-in-class drugs as well as orphan drugs have been found to encompass an increasing and meaningful portion of new drug approvals.[5−8] However, studies evaluating improvements in therapeutic benefit report declining innovation as most new drugs were found to only offer minor clinical advantages over existing treatments.[2]

Although these noncount based indicators can be used to highlight important advancements in the pharmaceutical industry, they may underestimate the rate of innovation

occurring as they focus on an outcome rather than the means to achieve a desired outcome. For example, measuring innovation using first-in-class or orphan drug designations will categorize all subsequent drugs in an area as "me-too" drugs, even if they are truly innovative. To more completely measure pharmaceutical innovation, we propose a new indicator of innovation for small molecule and peptide drugs based on structural novelty. This new indicator does not conflate innovativeness with the degree of success achieved as it is based on the structure of a NME at the time of its approval compared to the structures of prior FDA-approved NMEs.

Using our classification scheme based on structural novelty, we evaluated historical pharmaceutical innovation trends over the last 80 years and find that drug innovativeness has significantly increased over the last several decades. An important caveat is that our new indicator is not applicable to biologics which are complex in structure and are usually not fully characterized as they are generally derived from living

material. Biologics represent a vibrant area of research. However, small molecules have historically represented a majority of approved new therapeutic drugs and continue to be the dominant drug modality of new drugs.

## ■ METHODS

Since rings are the fundamental building blocks in the design of most small molecule drugs,[9] our new approach to characterizing drug innovation based on structural novelty relies on the concept of a molecular framework. The framework of a chemical structure is defined as the substructure consisting of all ring systems and all the chain fragments connecting them.[10] Therefore, the framework of a drug molecule can be thought of as the substructure that holds the side chains in place. Even though they ignore acyclic side chains, frameworks are core to the physical structure of molecules and can highly influence molecular properties and associated biological activity. These traits have made frameworks one of the most frequently applied concepts in medicinal chemistry.[11]

For this study, we use the framework at two levels of structural information. This is illustrated in Figure 1 using the
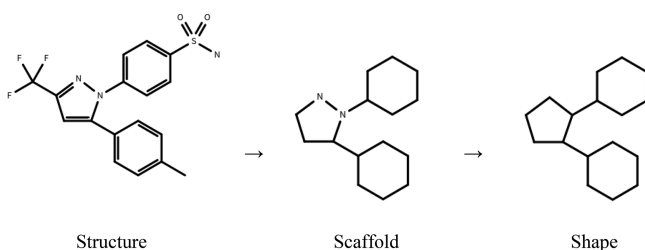


Figure 1. Extraction of a scaffold and shape from a drug structure.

anti-inflammatory drug CELEBREX (celecoxib). Frameworks at the scaffold and shape level only describe a molecular structure's topology (i.e., atom–atom connectivity), as they do not contain any information about stereochemistry or three-dimensional shape.

Pruning all acyclic side chains from this drug structure and removing all information about bond order yields what we will call the drug's scaffold. Removing from the scaffold all information about element types yields what we will call the scaffold's shape.

As the primary source of the major pharmaceutical advances over the last several decades,[12] the scope of our analysis was limited to NMEs. We compiled a data set of NMEs of interest by identifying the drugs containing one or more NMEs approved by the FDA's Center for Drug Evaluation and Research (CDER) as given in the Drugs@FDA database. The Drugs@FDA database contains information about most of the drugs approved by the FDA for human use in the United States since 1938.[13] An NME is a drug that contains an active moiety that has never been approved by the FDA either as a single ingredient drug or as part of a combination product.[14] We excluded diagnostic and imaging agents in order to focus our analysis on NMEs used to treat or prevent serious medical conditions. Only organic NMEs including small molecules and peptides that have a framework (cyclic compounds) were included in the final data set. The final data set consisting of 1,089 organic, cyclic NMEs and their structural innovation classifications is provided in the Supporting Information.

All of the NMEs in the final data set are included in the CAS REGISTRY, a comprehensive and authoritative database of chemical compounds reported in the scientific literature dating back more than 150 years. The CAS REGISTRY was used to obtain information about the frameworks of each NME in the final data set. This information was extracted from the framework data that CAS algorithmically extracts and systematically stores for every registered substance that meets certain criteria.[15]

The CAS REGISTRY was also used to obtain insights on the origins of the NMEs by identifying compounds that share shapes and scaffolds with the NMEs. The year an Investigational New Drug application (IND) for a drug was submitted in the United States was also used to obtain insights on the origins of the NMEs by serving as a reasonable proxy for the start of clinical development. FDA's Drugs@FDA database and the Federal Register[16] were used to collect IND years for each drug containing a NME of interest.

Because we are interested in contributions to early stage drug discovery, the organization that discovered the NME, rather than the organization that developed it or secured drug approval, was credited with the innovation. We made this determination using the AdisInsight "originators" classification; the originator usually refers to the institution(s) that AdisInsight reviewers concluded had originally invented or discovered the active ingredient of a given drug.[17] To account for mergers and acquisitions, we only credited the acquiring organization with NMEs that had an IND year after the year of an acquisition. For example, only NMEs originated by Wyeth with an IND year after the 2009 acquisition of Wyeth by Pfizer were credited to Pfizer.

## ■ RESULTS AND DISCUSSION

Innovation is the act of making something different that generates value and ranges from changing something that already existed (incremental advances) to creating something that did not previously exist (major breakthroughs). The use of frameworks as the basis for analysis has certain limitations as it fails to capture other ways in which pharmaceutical companies innovate, including, for example, incremental innovations that may occur through the modification of acyclic side chains. This approach also does not distinguish between incremental innovations associated with ring or linker modifications as different frameworks might actually describe very similar structures whose shapes consist of ring sizes or linker lengths that are different by only one atom or whose scaffolds only differ by a single heteroatom position. Nevertheless, frameworks are a conceptually simple way to represent the core structures of NMEs and serve as an insightful unit of analysis.

Our approach to characterizing structural innovation among drugs is based on a classification scheme that takes into account molecular framework information (at the scaffold and shape level) and approval year information. Each NME is assigned to one of three classes: Pioneers, Settlers, and Colonists, as defined in Table 1.

Each framework can be thought of as a region of the vast chemical space consisting of structurally similar molecules. Pioneers, the more innovative category for the purposes of this study, are the first drugs approved in previously unoccupied territories (shapes that have never been used as the basis for a drug). For a given shape, Pioneers are occasionally (just over 22% of the time) followed by one or more Settlers that have the same shape but different scaffold as the Pioneer or one or

**Table 1. Definitions for Three Classes of NMEs Based on Their Shape and Scaffold**

| Class | Definition | Total Count |
|---|---|---|
| Pioneer | A NME whose shape and scaffold were not used in any previously approved drug. | 511 |
| Settler | A NME whose shape was previously used but its scaffold was not used in any previously approved drug. | 201 |
| Colonist | A NME whose shape and scaffold were used in a previously approved drug. | 377 |

more Colonists that have the same shape and scaffold as the Pioneer. Similarly, Settlers are occasionally (just over 21% of the time) followed by one or more Colonists that have the same shape and scaffold as the Settler. Since innovation ranges from incremental advances to major breakthroughs, all three classes contribute to important advancements in the pharmaceutical industry.

**Trends in Structural Innovation.** After 35 years of a relatively consistent mix, there has been a significant shift in the composition of the approved NMEs on an absolute basis over the last 15 years. The behavior of the three classes over time is shown in Figure 2a. Between 1970 and 2005, Colonists were the largest class which might be a reflection of a widely



(a)



(b)
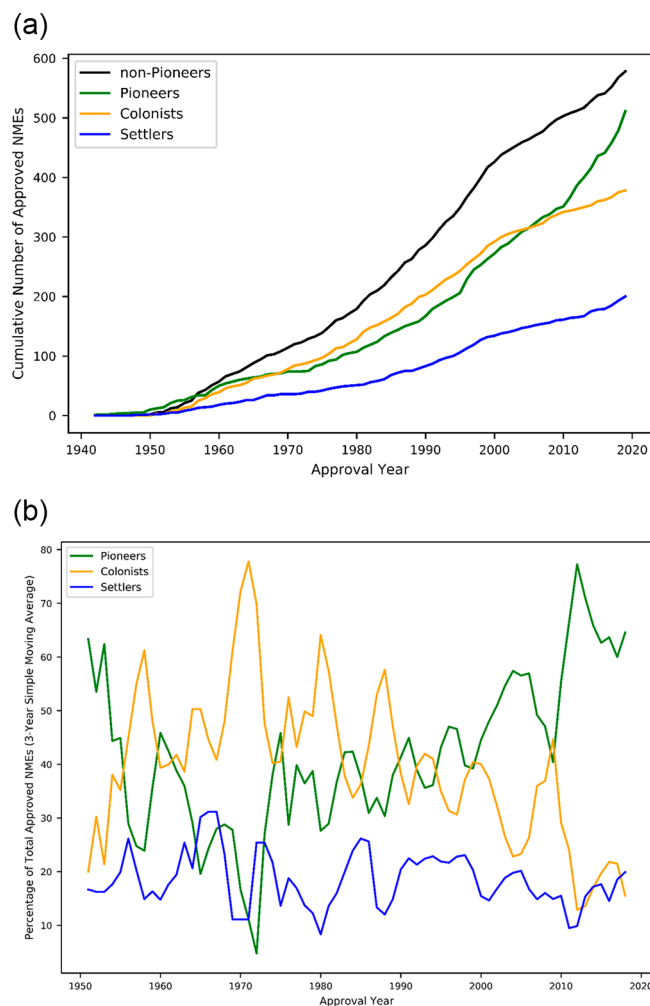
**Figure 2.** (a) Cumulative distribution plots for the three classes of NMEs. (b) 3-year moving average plots for the three classes of NMEs.

held assumption that it is only worth hunting for drugs in known, drug-rich regions of chemical space. This idea was reinforced by Sir James W. Black, the winner of the 1988 Nobel Prize in medicine, who famously stated that "The most fruitful basis for the discovery of a new drug is to start with an old drug".[18]

The rate of growth in Pioneers increased significantly between 1990 and 2019. The adoption of new synthetic and screening methods and the creation of screening libraries with greater diversity are possible contributing factors to this change.[19] A rough extrapolation of the Pioneer curve (green) and the non-Pioneer curve (black), consisting of both Settlers and Colonists, seems to show that (if the current trends continue) these curves will intersect within the next 10 years, at which point half of all NMEs will be based on a shape that was not used in any previous drug at the time of approval.

In order to identify underlying trends on a relative basis, a moving-average over a 3-year time period was utilized to smooth the data so as to reduce the effect of year-to-year fluctuations. As seen in Figure 2b, there has also been a significant shift in the composition of the approved NMEs on a relative basis over the last 20 years. Between 1980 and 2000, the mix of both Pioneers and Colonists as a percentage of the total approved NMEs mostly fluctuated between 30% and 50%. However, over the last 20 years, the mix of Pioneers has increased and fluctuated between 60% and 80% over the past decade, while the mix of Colonists has decreased and only fluctuated between 10% and 25% over the past decade.

Given Pioneers were the dominant source of structural innovation over the last several decades and the significant increase in Pioneers over the past decade, we focused the remaining analysis on the 248 Pioneers approved over the last 20 years. Before exploring where these structural innovations are being found and who discovered them, we first assess the therapeutic value associated with Pioneers. This assessment also allows us to examine the robustness of our results with respect to an alternative measure of innovativeness, specifically, a drug's clinical advantages over existing treatments.

**Therapeutic Benefits of Structural Innovation.** In order to examine the clinical advantages of Pioneers over existing treatments (therapeutic value), we utilized the breakthrough therapy designation as a proxy of a drug's ultimate clinical impact. The breakthrough therapy designation was introduced as part of the FDA Safety and Innovation Act of 2012 to help shorten the development and review time of promising new therapies intended to treat a serious or life-threatening disease for which there is unmet medical need and for which there is preliminary clinical evidence to demonstrate a potential substantial improvement on a clinically significant end point compared with other available therapies. Drugs granted breakthrough status were deemed to be therapeutically innovative.

Between 2013 and 2019, 73 (26%) of the 276 new therapeutic drugs approved by the FDA's CDER were granted breakthrough drug status. About 56% (41) of these were organic, cyclic drugs, almost 83% of which (34) included at least one Pioneer. Taking a deeper look, just over 28% of the drugs which included at least one Pioneer were judged as therapeutically innovative compared to nearly 11% of the drugs which only included non-Pioneers (Settlers and Colonists). As a result, we find Pioneers are significantly (2.6 times) more likely to be the source of promising new therapies. This

suggests our new indicator of innovation based on structural novelty has sensible properties.

**Origins of Structural Innovation.** By utilizing the CAS REGISTRY, we are able to determine the "age" as well as the "population" of the scaffolds for the Pioneers. Our definition of "age" is the number of years between the year of first report in the CAS REGISTRY for the scaffold of a Pioneer and the year in which the IND for a drug containing a Pioneer was submitted in the United States. The distribution of these scaffold ages is shown in Figure 3. This distribution shows that most of the scaffolds of Pioneers were reported in scientific literature within a few years of the IND of a drug.
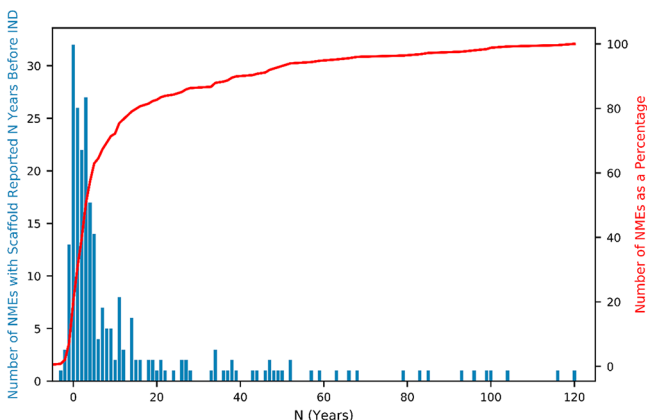


**Figure 3.** Distribution of scaffold ages for Pioneers.

The plot in Figure 3 shows that the scaffolds of 63% of the Pioneers were reported five years or less prior to the IND of the drug. Looking more closely, we find that this percentage has been increasing over the last 20 years. This trend is shown in Table 2, where the Pioneers have been bucketed into five-

**Table 2. Growth in the Percentage of Pioneers on New Scaffolds**

| Time Interval | Percentage of Pioneers with a scaffold reported five years or less prior to IND |
|---|---|
| 2000−2004 | 50% |
| 2005−2009 | 55% |
| 2010−2014 | 63% |
| 2015−2019 | 72% |

year blocks according to their approval year. This suggests new scaffolds (scaffolds that were first reported in the CAS REGISTRY five years or less prior to the IND year) are now significantly more likely to be the source of a Pioneer.

Another interesting origin-based insight is the "population" of the scaffolds for the Pioneers. Our definition of "population" is the number of structures in the CAS REGISTRY that shared the scaffold of a Pioneer at the time the IND for that drug was submitted in the United States. The distribution of these scaffold populations is shown in Figure 4. This distribution shows that most of the scaffolds of Pioneers were on less populated scaffolds.

The plot in Figure 4 shows that the scaffolds of 67% of the Pioneers were populated with 50 or less other compounds at the time of IND. After bucketing the Pioneers into five-year blocks according to their approval year, this percentage has been relatively consistent over the last 20 years. This seems
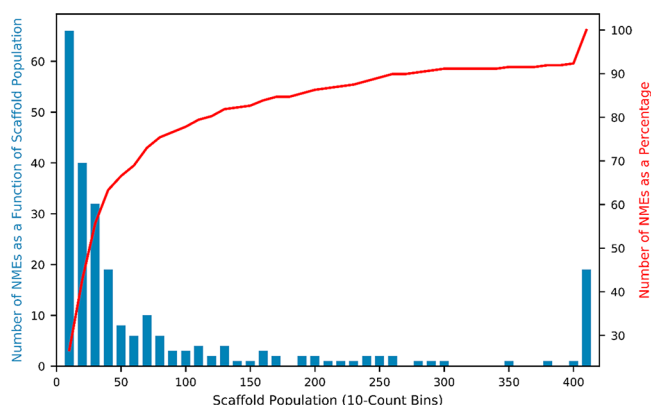


**Figure 4.** Distribution of scaffold populations for Pioneers. The last bar in this figure represents the (19) scaffolds populated with over 400 compounds.

logical given the relatively young age of the scaffolds as earlier discussed. However, 46% of the Pioneers on scaffolds over 5 years old were populated with 50 or less other compounds compared to 79% of the Pioneers on scaffolds 5 years old or younger. This suggests, regardless of the age of a scaffold, less populated scaffolds tend to be the source of a Pioneer.

**Originators of Structural Innovation.** In order to examine where structural innovation is more likely to arise within the pharmaceutical ecosystem, we divided the Pioneers into two groups: those originating in organizations that we define as "Big Pharma" and those originating in other organizations (Rest of Ecosystem or ROE). Our data may underrepresent the contribution by small companies in cases where a large company acquires a small company early in the discovery phase, but we believe this to be a minor concern given we attempted to credit the organization(s) that originally invented or discovered the NME with the innovation. While some of the organizations included in Rest of Ecosystem could be considered large for other purposes outside of our study, our analysis used the following 12 companies to constitute Big Pharma: AbbVie, AstraZeneca, Bayer, Bristol-Myers Squibb, Eli Lilly, GlaxoSmithKline, Johnson & Johnson, Merck & Co., Novartis, Pfizer, Roche, and Sanofi. These organizations were listed as the largest pharmaceutical companies over the 1995−2005 period[20] and remained some of the largest pharmaceutical companies over the last two decades based on drug-related revenue.

As seen in Figure 5, Big Pharma accounted for 43% of the Pioneers in 2000−2004. However, this percentage has gradually decreased over the three subsequent five-year time periods to 30% despite an increase in the number of Big Pharma originated Pioneers over the past decade. This suggests the significant increase in the number of ROE originated Pioneers relative to the number of Big Pharma originated Pioneers was the dominant factor influencing this decline.

It appears that both Big Pharma and the other organizations in the pharmaceutical ecosystem are finding more and more Pioneers on new scaffolds (scaffolds that were first reported in the CAS REGISTRY five years or less prior to the IND year) as shown in Figure 6.

This may suggest that when it comes to innovation, scale is no longer as much of an advantage as it once was and that it might become even less so in the future as the basis of competition is being redefined. Given the increasing
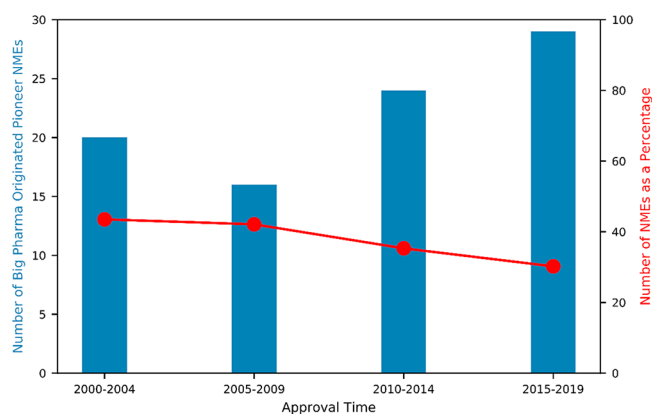
**Figure 5.** Pioneers originated by Big Pharma as a function of approval year.
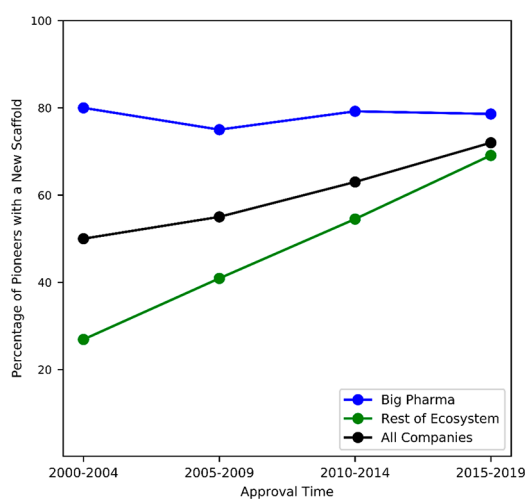


**Figure 6.** Pioneers with a new scaffold as a percentage of all Pioneers originated by different sectors of the pharmaceutical ecosystem.

importance of in-silico methods in the drug discovery process and the wider availability of high quality compound libraries, success increasingly hinges on the ability to derive better insights from data rather than acquiring the raw data itself. This may be a contributing factor to the pharmaceutical industry's improving R&D productivity, as measured by the average R&D cost per new approved drug.[21,22]

## CONCLUSION

Despite recent concerns over an innovation crisis, this analysis shows pharmaceutical innovation has actually increased over the last several decades based on the structural novelty of approved NMEs. The higher proportion of Pioneers over the most recent decade is a sign that innovation within the industry is accelerating rather than slowing. It is also an encouraging sign for the state of innovation in drug discovery that these Pioneers are significantly more likely to be the source of promising new therapies that are expected to provide substantial clinical advantages over existing treatments. Drug hunters are discovering Pioneers in newer and less explored regions of chemical space as they are increasingly found on scaffolds first reported in the CAS REGISTRY five or less years prior to their IND year or on scaffolds populated with 50 or less other compounds at the time of IND.

As scale becomes less of a strategic advantage, Big Pharma's share of Pioneers has decreased even though the number of Big Pharma originated Pioneers has increased. This has created a structural innovation gap between Big Pharma and the Rest of Ecosystem which has widened over the last two decades as the Rest of Ecosystem is now responsible for originating almost 3 out of every 4 Pioneers. Pioneers originated by the Rest of Ecosystem are increasingly on new scaffolds, while a majority of Big Pharma originated Pioneers have historically been on new scaffolds.

The work presented here was intended as a study of drug innovation at a macro level. As a result, it included substances of various sizes with different degrees of complexity belonging to a range of functional and drug classes. Even though it was outside the scope of the present work to study specific subsets, such focused studies could yield additional insights into how innovation at a more micro level has changed over time. Other interesting subsets of our data set are the shapes and scaffolds of the Settlers and Colonists. Many of these shapes and scaffolds are privileged in the sense that they are seemingly capable of serving as ligands for a diverse array of target proteins. A separate study of the Settlers and Colonists as well as their side chains could provide insights into possible target-specific innovation trends.

As it often takes more than 10 years after initial discovery for an experimental drug to gain FDA approval, any measure of drug innovation that relies on the time of approval incorporates a significant time lag between initial discovery and ultimate approval. However, characterizing drug innovation based on structural novelty provides a means to assess the forward-looking innovation potential of an experimental drug at the time of initial discovery by comparing its framework information (at the scaffold and shape level) with prior FDA-approved drugs. Therefore, a separate study of drug candidates with publically disclosed structures currently in clinical development could provide additional insights into innovation trends at an FDA regulatory review level and serve as a leading indicator of innovation trends at an FDA approval level.

Given the tremendous opportunity represented by the vast amount of chemical space yet to be explored, drug-hunters of all types will continue pushing the boundaries to find promising new therapies in previously unexplored areas of chemical space. The race to discover these new drugs will be fueled by further advancements in screening approaches and in-silico methods (including innovations related to machine learning algorithms and molecular representations). However, comprehensive data on known shapes and scaffolds can fast track the identification of meaningful open areas of chemical space (shapes or scaffolds that are potentially important but have never been used as the basis for a molecule) to further explore.

## ASSOCIATED CONTENT

**ⓢ Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsmedchemlett.0c00319.

> List of all of the NMEs in the final data set and their structural innovation classifications (XLSX)

# ■ AUTHOR INFORMATION

## Corresponding Author

**Todd J. Wills** − *CAS, Columbus, Ohio 43210-0012, United States;* ◉ orcid.org/0000-0003-3407-342X; Email: twills@cas.org

## Author

**Alan H. Lipkus** − *CAS, Columbus, Ohio 43210-0012, United States;* ◉ orcid.org/0000-0001-9449-3915

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsmedchemlett.0c00319

## Author Contributions

Both authors shaped the concept, designed the analysis, interpreted the results, wrote, reviewed, and edited the manuscript, and read and approved the final manuscript.

## Notes

# ■ REFERENCES

(1) Munos, B. Lessons from 60 years of pharmaceutical innovation. *Nat. Rev. Drug Discovery* **2009**, *8*, 959−968.

(2) Naci, H.; Carter, A. W.; Mossialos, E. Why the drug development pipeline is not delivering better medicines. *Br. Med. J.* **2015**, *351*, h5542.

(3) Kesselheim, A. S.; Wang, B.; Avorn, J. Defining "innovativeness" in drug development: a systematic review. *Clin. Pharmacol. Ther.* **2013**, *94*, 336−348.

(4) Mullard, A. 2019 FDA drug approvals. *Nat. Rev. Drug Discovery* **2020**, *19*, 79−84.

(5) Lanthier, M.; Miller, K. L.; Nardinelli, C.; Woodcock, J. An improving approach to measuring drug innovation finds steady rates of first-in-class pharmaceuticals, 1987−2011. *Health Aff.* **2013**, *32*, 1433−1439.

(6) Phillips, M. A.; Lee, J. R. Recent orphan drugs that are first-in class. *Expert Opin. Orphan Drugs* **2014**, *2*, 759−763.

(7) Attwood, M. M.; Rask-Andersen, M.; Schiöth, H. B. Orphan Drugs and Their Impact on Pharmaceutical Development. *Trends Pharmacol. Sci.* **2018**, *39*, 525−535.

(8) Woodcock, J. Innovation in New Drug Approvals of 2019 Advances Patient Care Across a Broad Range of Diseases. *FDA Voices* **2020** https://www.fda.gov/news-events/fda-voices/innovation-new-drug-approvals-2019-advances-patient-care-across-broad-range-diseases (accessed on August 18, 2020).

(9) Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845−5859.

(10) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.

(11) Hu, Y.; Stumpfe, D.; Bajorath, J. Computational Exploration of Molecular Scaffolds in Medicinal Chemistry. *J. Med. Chem.* **2016**, *59*, 4062−4076.

(12) Kesselheim, A.; Avorn, J. The most transformative drugs of the past 25 years: a survey of physicians. *Nat. Rev. Drug Discovery* **2013**, *12*, 425−431.

(13) *Drugs@FDA: FDA-Approved Drugs.* https://www.fda.gov/drugsatfda (accessed on August 18, 2020).

(14) Branch, S. K.; Agranat, I. "New Drug" Designations for New Therapeutic Entities: New Active Substance, New Chemical Entity, New Biological Entity, New Molecular Entity. *J. Med. Chem.* **2014**, *57*, 8729−8765.

(15) Lipkus, A. H.; Watkins, S. P.; Gengras, K.; McBride, M. J.; Wills, T. J. Recent Changes in the Scaffold Diversity of Organic Chemistry As Seen in the CAS Registry. *J. Org. Chem.* **2019**, *84*, 13948−13956.

(16) *Federal Register.* https://www.federalregister.gov (accessed on August 18, 2020).

(17) *AdisInsight.* https://adisinsight.springer.com (accessed on August 18, 2020).

(18) Raju, T. N. The Nobel Chronicles. *Lancet* **2000**, *355*, 1022.

(19) Liu, R.; Li, X.; Lam, K. S. Combinatorial Chemistry in Drug Discovery. *Curr. Opin. Chem. Biol.* **2017**, *38*, 117−126.

(20) Gautam, A.; Pan, X. The changing model of big pharma: impact of key trends. *Drug Discovery Today* **2016**, *21*, 379−384.

(21) Bloom, N.; Jones, C. I.; Reenen, J. V.; Webb, M. Are Ideas Getting Harder to Find? *Am. Econ. Rev.* **2020**, *110*, 1104−44.

(22) Ringel, M. S.; Scannell, J. W.; Baedeker, M.; Schulze, U. Breaking Eroom's Law. *Nat. Rev. Drug Discovery*, published online April 16, 2020; **2020**, DOI: 10.1038/d41573-020-00059-3.