Measuring the Relationship Between Innovative Drugs and Adverse Event Data

Jonathan Bryan

March 16, 2015

Abstract

A key discussion among regulatory and industry stakeholders has been whether the drugs approved by the FDA are not only safe and effective but also innovative. Innovation is now at a premium, set against the rising cost of health care and the diminishing returns of research and development expenditure by biopharmaceutical stakeholders. Despite the need to discover, develop and approve new high-quality drugs, the metrics and models that help researchers identify innovative drugs are either insufficient or nonexistent. A paper that classified drug innovation using review designations of drugs approved from 1987 to 2011 was used and tested against the adverse event data found within the FDA Adverse Event Reporting System. Total adverse events, patient demographics, and outcomes were recorded for each drug. It was determined that the incidence of adverse events of drugs approved by the FDA may follow a power law with unstable variation and capacity for "black swan" behavior. In addition, logistic regression yielded a weak relationship between "advanced-in-class" drugs and "addition-to class" drugs except at the fat tails.

Background

Challenges to Developing and Measuring Innovative Drugs

Every year the U.S. Food and Drug Administration (FDA) reviews and approves drug applications on the basis of safety and effectiveness. This regulatory review system supports the engine of the multibillion dollar biopharmaceutical industry. A key discussion among regulatory and industry stakeholders has been whether the drugs approved by the FDA are not only safe and effective but also innovative. Innovation is now at a premium, set against the rising cost of health care that has pushed policymakers and health payers to embrace a value-oriented paradigm. To compound this issue, the cost of developing new drugs, and not just innovative drugs, has continued to increase, following a daunting trend known as Eroom's law (Moore's law backwards).

Despite the need to discover, develop and approve new high-quality drugs, the metrics and models that help researchers identify innovative drugs are either insufficient or nonexistent. Traditional efforts to quantify innovation have looked at the "frontend" of the drug innovation: drug development. These metrics include industry financial data such as research and development expenditures, venture capital investments, as well as regulatory data such as FDA approval rates of new drugs, novel therapeutic mechanisms, and the utilization of expedited approval pathways. Only recently have "backend" or outcome metrics been gaining attention, a function of the increase in data availability and the need for quantifying drug quality for real-world applications. Such outcome metrics include adverse event reports, economic evaluations and quality-adjusted life years. While each of these "frontend" and "backend" metrics provide critical information about the velocity and nature of drug innovation, many are used in isolation or suffer from gaps in data and fidelity. Moving forward, stakeholders will be seeking to aggregate disparate sources of drug information to formulate a more complete picture of drug innovation. Ascertaining which data sources will provide the strongest "signal" for innovation will be an active project.

Lanthier et al.

A seminal paper by researchers at the FDA presented a useful criterion for assessing the innovative status of drugs based upon their review status. They came up with three distinct classes as show in Table 1.[1]

Table 1.

Lanthier Drug Innovation Classes

First-in-class	Pharmacologically innovative because each represents a new pathway for treating a disease.
Advanced-in-class	Drugs not first-in-class but received a priority review designation, which is reserved for medicines that potentially offer major advances in treatment. Priority review designation suggests a measurable degree of innovation in the clinical potential of a drug.
Addition-to-class	Function similarly to other drugs in their class and do not offer substantial advantages in safety or efficacy over existing products. Generally represent a lower degree of innovation because at the time of FDA review, they did not distinguish themselves in terms of potential clinical benefit.

The paper was able to segment every drug approved since 1987 into each of these classes based upon their review designation and novelty. Review designations are given before a drug is approved for marketing and sale and thus their analysis constitutes a “backend” metric. This paper seeks to use this practical system for classifying innovative drugs and test it against the “frontend” metric of adverse events. Such analysis may furnish insight into how well FDA can determine innovative drugs before they are marketed.

Drug Safety, Adverse Events and Drug Innovation

Drug safety is a critical component of how new and approved drugs are reviewed and marketed for sale in the United States. In the pre-market phase that encompasses drug development, industry sponsors conduct numerous clinical trials and analysis to determine a safety and risk profile for the drugs they are developing. An important concept is that the safety of a drug is relative to its therapeutic benefit. Certain oncological drugs may expose patients to higher risks for adverse events but are deemed acceptable based upon the nature of the disease and the therapeutic benefit the drugs provide. During the premarket phase, industry sponsors will conduct test used to measure of toxicity, dosing, and side effects. After a drug is approved, a sponsor may be instructed by FDA to conduct postmarket studies to gather more evidence on a drug’s safety. FDA also engages in its own postmarket surveillance through multiple systems that gather data from voluntarily submitted adverse event reports, health claims, and electronic health records.

FDA Adverse Event Reporting System (FAERS)

The FDA Adverse Event Reporting System is an important tool that allows FDA to receive voluntary adverse event reports from multiple stakeholders. Stakeholders are able to electronically submit information on an adverse event and have the ability to identify important medical information such as suspected and concomitant drugs, demographics, and outcomes. FDA uses FAERS to spur inquiries into a drug’s safety profile that may lead to changes in a drugs marketing status, labeling information, medical use, and new public safety communication. FAERS holds over five million of such records and FDA has made available the entire database online for download and through its openFDA application programming interface (API) for developers.

Despite its availability and FDA stewardship, using FAERS outside of the paradigm of so-called hypothesis generating studies can be problematic. As previously mentioned, all reporting through FAERS is done on a voluntary basis and the majority of reports are submitted by drug manufacturers. This presents challenges related

[1] Lanthier, Michael, et al. "An improved approach to measuring drug innovation finds steady rates of first-in-class pharmaceuticals, 1987–2011." *Health Affairs*32.8 (2013): 1433-1439.

to incomplete data, data quality, and reporting bias. Certain drug classes may be reported more than others as well as certain brands within classes. In addition, reports to not require any standard terminology or codes needed to report specific drugs. Exacerbating this issue further is the lack of publically available and readily accessible data on specific drug utilization. Knowing how many drugs are in being used by patients is critical to understanding the context of adverse event data.

Measuring Drug Innovation with Adverse Event Data

Notwithstanding these limitations, the data contained within FAERS is helpful for two primary reasons: its size, and the diversity of data contained therein. For this paper, a model for drug innovation was developed that specified safety and effectiveness as the two primary factors that interact with each other to output the raw innovation "score" of a drug. Adverse event data was used as measure of a drug's safety for this model while no metric was used to measure effectiveness. A key assumption of this model is that the effectiveness of each drug within a certain drug class is equal while the comparative effectiveness between drug classes themselves is not. As an example, both "addition-to-class" and "advanced-in-class" statins can be thought of as having equal effectiveness while comparison between statins and non-steroidal anti-inflammatory drugs would not. This is a key assumption, done for purposes of model consistence and is recognized as being unrealistic in a real-world setting.

Another explicit assumption of the model is that the more innovative and the less novel a drug is the lower number of adverse events should be reported for that drug. These results in "advanced-in-class" drugs at the top of the innovation hierarchy, as "first-in-class" drugs run the risk of unforeseen effects given their novel characteristics and "addition-to-class" drugs offering no additional therapeutic and safety benefit. This assumption may be more reasonable than the first but could still fail under real-world circumstances such as minute but novel advancements in "advanced-in-class" drugs causing additional adverse events, and given their novel nature, higher burdens of safety and effectiveness of "first-in-class" drugs reducing the incidence of adverse events.

Hypothesis

Given the innovative nature of "advanced-in-class" drugs, as determined by Lanthier et al., such drugs should have a lower incidence of adverse events as reported by the FAERS.

Methods

The Lanthier data containing 645 drugs approved from 1987 to 2011 was obtained from the paper appendix. The FAERS data was downloaded from the FDA webpage and stored on a local system. Data processing and analysis was done using the Anaconda distribution of the Python programming language and the statistical libraries of Pandas, and scikitlearn. The FAERS database contained reports from 2004 to the second quarter of 2014. The FAERS database is segmented into multiple tables including ones for drug, demographic, and outcome information. Each table contains over five million reports and each report is linked to the various tables by an unique ISR code. These codes are not found within the Lanthier et al. dataset and thus a link needed to be established to assign specific adverse event reports to specific drugs. This paper used the drug trade name as that key. Each adverse event report table was searched using the trade name for the drugs from the Lanthier data set. Drug names were shortened to increase the capture of any name abbreviation found within the FAERS reports. As an example "Mefloquine Hydrochloride" may only be reported in an FAERS record as "Mefloquine". Unfortunately, the naming convention for drugs within the database various considerably and thus it is expected some reports were missed.

Each search recorded a total number of adverse event reports for each drug including total serious events, male patients, female patients, hospitalizations, resulting disabilities, congenital anomalies, life-threatening events, and deaths. The number of adverse events per year was also calculated based upon the year a drug was approved up to 2014. Additional pre-processing of the data included adjustments for the inflation of adverse events reported to FAERS since 2004. As shown in Figure 1, the number of adverse events as steadily grown over

the years, owing to a number of factors including increased pharmacovigilance and utilization. It was recognized that such growth may underweight the adverse events reported in closer to 2004 and overweight those reported in later years. Thus an adjusted adverse events and adjusted adverse events per year variables were calculated based upon 2004 adverse events. The resulting correction can be seen in Figure 2. Despite an overall dampening of total adverse events reported, the general structure of the data remained.

Figure 1

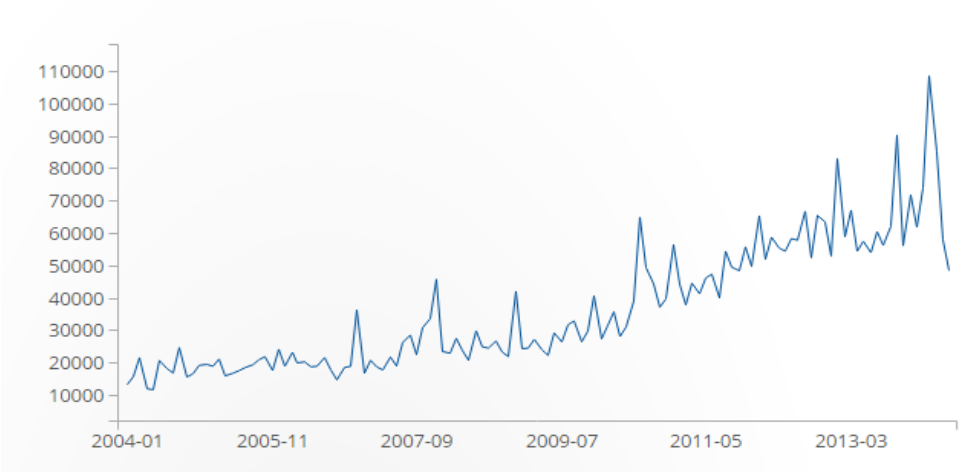
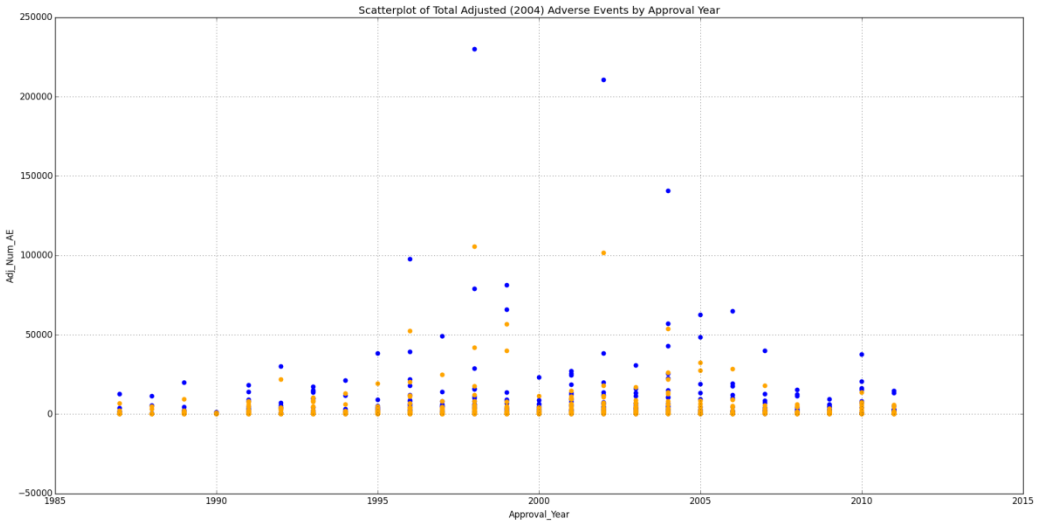


Figure 2



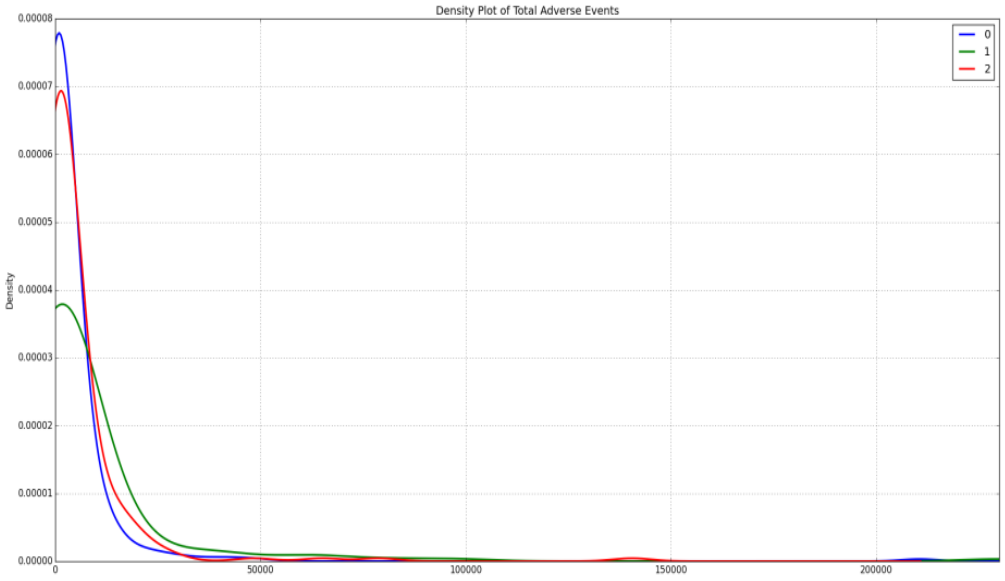
Before running any inferential analysis, summary level statistics were done and an empirical distribution was sought for the total adverse event data. It was determined that most of the independent variables were highly correlated with each other except for total adverse events, life threatening events, congenital anomalies, disabilities, deaths, and adjusted number of adverse events. Logistic regression analysis was then used to establish a relationship between these predictor features and the “addition-to-class” and “advanced-in-class” drugs. Recursive feature elimination was used to further narrow down the most optimal feature set.

Results

Adverse Events of Approved Drugs Follow a Power Law

A central finding of this paper was that the population of total number of adverse events, and indeed their associated outcomes, associated with each approved drug appears to follow a power law. This relationship is also consistent across each innovation class and more pronounced for “advanced-in-class” drugs (Figure 3). Using the Kolmorgorov-Smirnof one sided test it was found that this empirical distribution best fit a Pareto type II or Lomax distribution with support at x = 0 and a exponent value of a = 0.45. A key feature of power law distributions is that the exponent should preferably be above two to have a stable mean and above three to have a stable variation. Any exponent lower than three can display extreme value or what is also known as “black swan” behavior. If FAERS data does indeed indicate the possibility of black swan behavior, then this may have implications for the ability of FDA to capture and mitigate future drugs that could exhibit extreme incidences of adverse events.

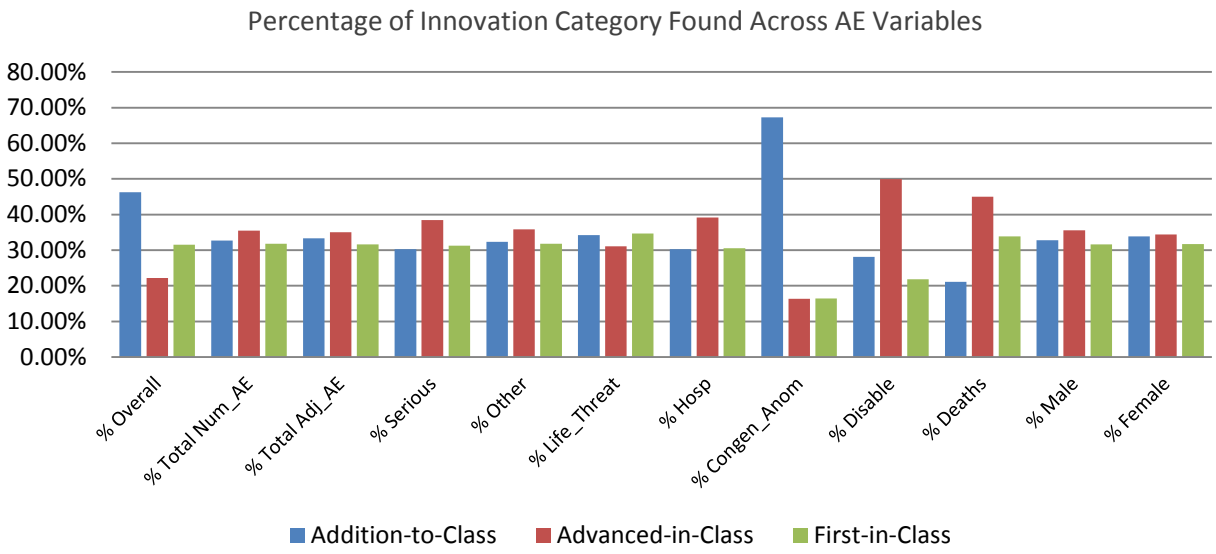
Figure 3



“Advanced-in-Class” Drugs Display a Disproportionate Amount of Adverse Events Reported

While representing only 22 percent of approved drugs since 1987, “advanced-in-class” drugs make up over 35 percent of total adverse event reported and over 45 percent over total disabilities and deaths reported to FAERS (Figure 4). However, “advanced-in-class” drugs represented only 31 percent of life threatening adverse events, lower that both “addition-to-class” (34 percent) and “first-in-class” (34 percent) complicating the original finding.

Figure 4



FAERS Data is a Weak Predictor of Drug Innovation

Logistic regression analysis yield mixed results that in sum provide very little evidence of a relationship between FAERS data and drug innovation. Individual logistic regression models were built for each of the most promising predictor variables as determined during the correlation analysis, including: total adverse events, life threatening events, congenital anomalies, disabilities, deaths, and adjusted number of adverse events. A comprehensive model was also built including all of the predictor variables and recursive feature elimination was used to determine an optimal model that included: total adverse events, life threatening events, deaths, number of males, number of females, and adjusted number of adverse events per year.

The null rate of "addition-to-class" drugs compared to "advanced-in-class" drugs was 67 percent and was used as a baseline to judge model performance. Each of the individual models performed near 69 percent accuracy, with specificity for the null or "addition-to-class" predictions around 97 percent and sensitivity for "advanced-in-class" predictions around 7 percent. The comprehensive and optimal models performed better than the individual models but had equal levels of performance to each other with 73 percent accuracy, 97 percent specificity and 19.5% sensitivity. Analysis of the increase in performance regarding the sensitivity of the comprehensive and optimal models found that their higher responsive near the fat tails provided better predictions and captured many more "advanced-in-class" drugs that presided there. Near the more populous end of the distribution, the two classes were near indistinguishable.

Conclusion

Implication of this paper's results include evidence that "advanced-in-class" drugs have a higher incidence of adverse events overall, and that individually exhibit many more extreme values than "addition-to-class" drugs. Such a result could be readily explained if "advanced-in-class" drugs have higher rates of utilization that "addition-to-class" drugs. This is likely given their likelihood of delivering increase clinically meaningful benefit. Indeed subsequent analysis of the "advanced-in-class" drugs found within the fat tails of the distribution included some of the most popular drugs used including the TNF inhibiting anti-inflammatory drug Humira and the now recalled non-steroidal anti-inflammatory drug Vioxx.

Additional findings include the need for better safety and outcome data to accurately assess how quickly and efficiently the United States is approving not only new but innovative drugs. Such metrics need to be bolstered with complementary utilization data to provide the proper context with which to analyze adverse event numbers.