A NEW CATALYST EVENT UNCOVERED FOR INVESTING IN THE PHARMACEUTICAL AND BIOTECH MARKETS

Executive Summary

Investing in pharmaceutical and biotech companies and predicting the catalyst events in the marketplace can be a challenging proposition without the proper alternative data and an understanding of its technical details. This paper explores the impact of a clinical trial start on a company's value. This newly discovered catalyst event has been proven to produce highly predictable outcomes versus the market with an accuracy of over 70%.

Furthermore, we describe a simple investment portfolio approach centered around this new catalyst event that delivered a 32% annualized return over the last two years, nearly double that of any comparable global biotech and pharmaceuticals index.

Problem Statement

"Chasing headlines" for catalyst events in the biotech and pharmaceutical markets is a common frustration of investing in these spaces. Predicting these headlines in advance appears to be the common goal, as well as mastering the complexities of these markets for a more informed investment strategy.

Much of clinical trial data is publicly available, and it can potentially provide a treasure trove of information about companies and the treatment of diseases. This data can be particularly powerful when aggregated across specific groupings, such as for a single company, group of companies or disease areas. Once aggregated and enhanced, this data can easily uncover emerging trends in a company's strengths and strategy, as well as trends in the treatment of a disease and the likelihood of the standard of care changing in the next few years for those patients.

Unfortunately, this data in its publicly available condition is not organized or captured in a manner to facilitate aggregation at the company or disease level. Company names, drug names and disease descriptions are all unstandardized. Neither the drug's primary mechanism of action (the biologic pathway in which the drug works) nor the drug's commercial approval status are required information.

To get the most out of this data, most investors spend large amounts of time and money retrieving it from its various sources and correcting any errors, omissions or repetitions. Typically, this process requires multiple data analysts for extracting, cleaning and storing the data in a ready format, and it delays the flow of this information into investment decision-making. Only after all this is done can the data be properly analyzed.

Finally, the conventional catalysts used by investors to identify opportunities in pharmaceutical stocks may have some limitations that make them unreliable models for evaluating the risk associated with investing in a pharmaceutical or biotech company.

Below is a breakdown of some of the common catalysts and their limitations as well as a description of a new catalyst uncovered during our analyses of these. It is this new catalyst – the impact of a clinical trial start – and how performance predictions can be applied from it for short-term investments that form the purpose of this paper.

Traditional Catalyst Events in Pharmaceutical and Biotech Investing and Alternative Data

As described above, the primary challenge for investing in the pharmaceutical and biotech companies is in predicting catalyst events in advance of the news cycle. Applying proven alternative data sources such as an enhanced clinical trial database can provide the critical information and trends necessary to predict the following catalyst events with regularity and clarity.

1. **Earnings Reports** – For larger pharmaceutical companies, the resulting impact of these reports is more balanced between the current sales and the company's pipeline. With a clean clinical trial data feed, a pipeline fitness report for every company can be available at your fingertips. When this data is aggregated from the clinical trial level as we provide in BEAM, every reported clinical trial outcome can be easily understood in the context of the impact to that company.
2. **Clinical Trial Results and Outcomes** – For smaller companies, the bulk of their earnings reports are focused on the clinical progression of their potential products. Depending on the number of products they have in development and other factors related to the disease areas of their products, these clinical outcomes can have a significant short- to medium-term impact on their stock price. This is one of the more difficult catalysts to predict, because we need to predict each of a) the likelihood of the trial's success, b) the impact of that trial relative to the marketplace, and the most challenging, c) when we expect to hear these reports. We provide all the raw information needed for these, as well as specific performance estimates relative the market for every trial with a greater than 60% accuracy in each of the last 10 years. Additionally, we are the only ones to provide a clinical trial outcome reporting date that is based on recruitment estimates and primary endpoint timing. Multiple investment strategies, both short and long, can be applied using this catalyst.
3. **Mergers & Acquisitions** – One of the greatest impacts to the value of a small- to mid-sized biotech company is being acquired. Pipeline fitness reports aggregated from clinical trial data can be applied both to larger companies making acquisitions as well as to smaller companies being acquired. With both in hand, any number of approaches can be taken for predicting the best company pairs in terms of complementary profiles and the likelihood of a deal transacting.
4. **Clinical Trial Starts –**This catalyst was discovered during the process of developing another catalyst prediction model related to the outcomes of clinical trials (item #2 above). In many cases, the scan across the clinical trial period for relative stock performance peaks found that these peaks often occurred very early in the clinical trial period (typically within the first 28 days but most often within the first 17 days). After further analysis, we determined that this phenomenon was quite predictable, and that the timing was much easier to predict than the timing for reported outcomes. Just like our predictions for the relative stock price impact of trial completions, our predictions of "outperform" and "underperform" for clinical trial starts have been more than 70% accurate for each of the last 10 years in back-testing. Additionally, when

the individual performance predictions were aggregated, we could apply them towards developing short-term (two-week) portfolios with a significant return against other pharmaceutical and biotech indices.

Predicting Relative Stock Performance of a Trial Start

Our data is updated daily with new clinical trials (as well as data updates to current ongoing trials) and then processed to provide a clean version of the data elements described above plus many more. Once processed, this data can be aggregated at the company level to provide a clear picture of every pharmaceutical company's development pipeline. For every clinical trial, we provide the original data as it exists in the clinical trial registry, as well as the additional processed data elements and model predictions. The process captures and stores not only every clinical trial but also every version of that trial. With an average of ten versions for each clinical trial this data is ideal for developing any number of machine learning approaches for outcome prediction. The potential use cases for this data in modeling, prediction and data reporting are boundless.

For our trial start catalyst predictions, our data contains roughly a half dozen new trial announcements every day with a company-based prediction of relative stock performance based on the trial start (RSPTS) and a confidence interval for that prediction. Our predictions model is trained using a subset of data from our database in combination with company stock information. A Grid Search meta-estimator from the Scikit-Learn library simulates all possible parameters in search of the best approach for our Random Forest Classifier model. Our models are updated regularly and catalogued in order to maintain a point in time reference. In general, the importance of features remains stable across models. In order of importance from greatest to least, the value-impacting data elements with the highest relevance to the model are length of the clinical trial, company conducting the trial, disease area, phase and company stock price trend the week prior to trial start.

Typically, most of these predictions trend towards "outperform" against the market and this outperformance has been the primary focus of our research to date. However, there is a strategy still to be explored with the underperformance predictions when used alone or in combination with other models. For now, we will focus on a simple portfolio investment strategy applied to the model's performance predictions to demonstrate its usefulness. We are certain that the readers of this article will be able to apply other more sophisticated strategies to the model predictions found in our BEAM data.

Portfolio Approach Using the Trial Start Catalyst (RSPTS)

It is possible for a single company to start multiple trials in any given week. The company can be assigned varied performance predictions for each trial start dependent on the value-impacting elements related to each trial. To arrive at a stable investment approach, we have created a simple two-week portfolio roll-over. Every two weeks, we review the prior two weeks and collect the trial start predictions that came in over that time. We then aggregate these by company and compute a "buy ratio" score by multiplying the total number of outperform predictions for that company by their average confidence level and subtracting from this the total number of underperform predictions weighted by their average confidence level.

This computed score allows us to sort the companies with new information over this period and determine both a minimum cutoff score for investment (in this case 0.5), as well as a relative weighting

for the companies in our portfolio. As an additional control, we also compute what a simple market cap-based weighting would be for the full list of companies with clinical trials publications to further isolate the impact of our predictions from any general announcement effects.

We compute and combine our portfolio's daily returns and compare them to a broad market index over the full period of the exercise. To mitigate endpoint sensitivity, we also look at two-week rolling returns for our strategy versus the various benchmarks. For the most optimal benchmark comparison in our back test, we developed an index that reflects similar small to large cap ratio and US based vs. non-US based pharmaceutical and biotech companies as our prediction portfolios.

Portfolio Results Using the Relative Performance of Clinical Trial Starts

Back-testing the approach described above using the appropriate "point-in-time" prediction models for each prediction period generated an annualized return of 32% over a period of 24 months (from August 2019 to mid-September 2021). By choosing only companies with a computed "buy ratio" of 0.5 or higher and weighting them proportionally to develop the Selective Portfolio, our return was twice what it would have been had we simply built a cap-weighted portfolio across all companies with a clinical trial start (16%). More importantly, the selective technique delivered a better return versus our hybrid Blended Pharm/Biotech Index 15% or XPH, the S&P Select Industry Pharmaceuticals Index ETF, with its 19% annualized return.

	Total Multiple	Annual Return	Average Monthly Return	Average of Rolling 2W Returns	% of Rolling 2W Periods Selective Portfolio Outperforms
Selective Portfolio	1.77	32.3%	2.6%	1.2%	N/A
ALL Clinical Trial Starts Portfolio	1.34	15.7%	1.5%	1.0%	58%
Blended Pharm + Biotech Index	1.33	14.9%	1.5%	1.0%	57%

Table 1. Results of our Selective Portfolio based on our prediction model for clinical trial starts as compared to all clinical trial starts and to a comparable index of biotech and pharmaceutical companies overall

In addition, when looking at rolling returns, the selective technique outperformed both benchmarks in roughly 60% of all 2-week subperiods. These encouraging results over the last two years demonstrate that our selective portfolio approach drives short- and long-term results that outperform market indices during both good times and bad.

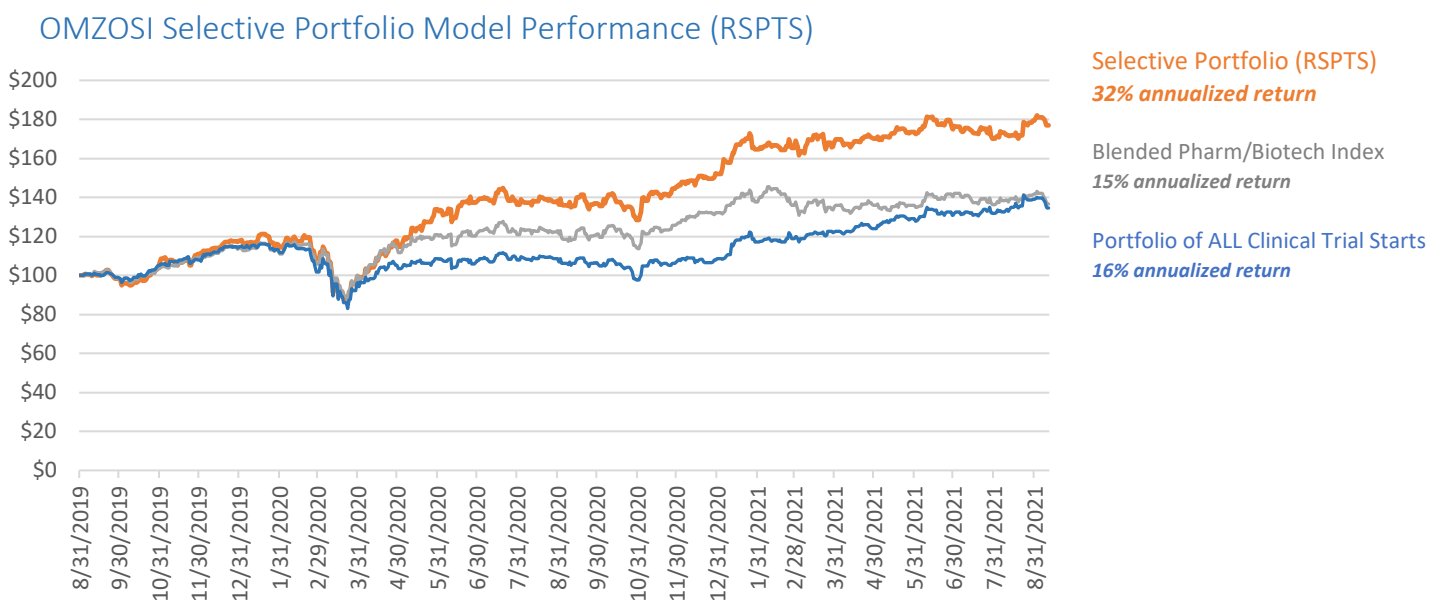

Chart 1. Return over time of our Selective Portfolio based on our prediction model for clinical trial starts as compared to all clinical trial starts and to a comparable index of biotech and pharmaceutical companies overall

Conclusion

While there is certainly a good deal of noise and external influences on stock prices for pharmaceutical and biotech companies, there are also several clear and predictable catalyst events that occur with reliable frequency when applying the appropriate alternative data for this marketplace.

The results outlined here could be further improved with a more sophisticated approach to the hold times for the companies in these portfolios. For the purposes of this paper, we focused on the selection of the portfolio with a more standardized exit strategy so that any resulting value impact could be attributed only to the company selection process related to the clinical trial start catalyst. This should be encouraging news for any investors looking to build portfolios that will outperform against the pharmaceutical and biotech markets.

About OZMOSI

OZMOSI is a consulting and data services firm founded in 2013 with the singular mission to provide pharmaceutical and biotechnology companies the support they need to make hard decisions in a high-risk and highly uncertain environment.

Combining expert pharmaceutical experience with advanced analytics, OZMOSI offers solutions supporting strategic clinical decisions. We provide state-of-the-art modeling, machine learning and visualization tools, illuminating the competitive landscape without sacrificing any of the details.

Applying our depth of industry knowledge and analytical skills the OZMOSI team has developed the most robust sets of data available for building prediction models and visualizations for our data clients in the pharmaceutical and investment industries.

About The Data:

BEAM Biotech and Pharmaceutical Company R&D Data contains over 130,000 industry-sponsored clinical trials, covering more than 1,000,000 point-in-time data versions, across 6,000 companies with over 1,800 mapped to ticker symbols. The robust point-in-time R&D data includes cleaned company names, drug names, Mechanism of Action, drug target information, whether a treatment is FDA approved or novel, disease area tags, oncology biomarkers and various text rich data elements for NLP development of categories and indicators. Additionally, the data provides performance prediction flags to better predict clinical trial success. Utilizing machine learning, the data offers elements to predict stock price performance related to trial completion and stock price performance within the first 28 days after the start of a trial.

The raw data is web-scraped/web-crawled from publicly available sources, including but not limited to clinicaltrials.gov. The data is processed and augmented using proprietary industry libraries.