ALY6980 – Capstone

06/30/2024

Prof. Roy Wada

Title: Unveiling Trends of Epilepsy Project

Assignment 12 – Individual Project Proposal

Group 7

Submitted By: Manav Jatakia

**Introduction:**

In the landscape of rare pediatric epilepsies, my project endeavors to harness the power of publicly available data and recent literature on epilepsy research to derive actionable insights. By focusing on this specific indication, the project aims to provide a comprehensive understanding of key factors influencing strategic decision-making for the sponsor's startup company. Through meticulous analysis of datasets and literature, I seek to contribute invaluable information regarding indication selection, clinical trial outcomes, pricing data, product-market fit, and dynamic market trends.

At the core of this individual endeavor lies the objective to synthesize insights from recent literature on epilepsy research, particularly emphasizing pediatric epilepsies. By delving into scholarly articles, studies, and research findings, the project aims to uncover emerging trends, innovative treatments, and gaps in understanding specific to rare pediatric epilepsies. Through this exploration, I intend to equip the sponsor's company with a nuanced understanding of the current landscape, enabling informed decision-making and strategic planning.

I will use publicly available data and databases to extract actionable insights tailored to the unique challenges and opportunities of rare pediatric epilepsies. Through rigorous data analysis, I aim to uncover patterns, trends, and correlations that can inform strategic initiatives, including indication prioritization, clinical trial design optimization, pricing strategies, and market positioning. By focusing on this specific indication, the project seeks to provide targeted recommendations that align with the company's mission to address unmet needs in pediatric epilepsy treatment.

The article "An Overview of Machine Learning Techniques for Detecting Pediatric Epilepsy" (Zurdo-Tabernero et al., 2023) offers valuable insights for the company's project focusing on rare pediatric epilepsies. Firstly, it provides a thorough understanding of machine learning methods tailored for identifying pediatric epilepsy, enabling the company to apply these techniques to analyze data and recognize patterns from publicly available datasets. This aids in determining which rare pediatric epilepsies to prioritize based on factors like prevalence, severity, and market demand. Additionally, the article evaluates the accuracy of different machine learning algorithms in seizure detection, assisting in predicting potential outcomes of clinical trials and guiding decisions on trial design and patient recruitment. Furthermore, the insights provided on EEG signal processing, feature extraction, and dataset characteristics can help the company extract pertinent pricing data, evaluate product-market fit, and understand market dynamics related to rare pediatric epilepsies. In summary, the article serves as a valuable guide for the company's strategic planning, influencing decisions regarding both current and future programs concerning rare pediatric epilepsies.

**Methodology:**

I will be utilizing the NIH RePORTER website for our project's research methodology. This invaluable resource has empowered us to delve deeply into the intricacies of rare pediatric epilepsies, allowing us to make well-informed decisions that drive our strategic direction. By meticulously analyzing the extensive repository of funded research projects, clinical trials, and grant awards, we have been able to pinpoint crucial areas of unmet medical need, identify potential collaborators, and track emerging trends within the epilepsy research landscape.

With a dataset comprising 11,964 rows and 52 columns, the wealth of information contained within offers a comprehensive insight into funding allocations and total project costs across various initiatives. Beyond financial details, the dataset encompasses crucial metadata such as country of origin, sponsoring entity, project commencement and completion dates, among others. These diverse data points serve as foundational pillars for analysis, providing valuable context and granularity essential for deriving meaningful insights. By delving into this dataset, we can uncover trends, patterns, and correlations that shed light on funding dynamics, project longevity, geographic distribution, and sponsor preferences.

For my project, I will adopt a similar approach to Michael Privitera's methodology in his article “Large Clinical Trials in Epilepsy: Funding by the NIH versus pharmaceutical industry.” (Privitera, 2006). Drawing upon peer-reviewed literature and leveraging insights from previous trials, I will gather and analyze data pertinent to rare pediatric epilepsies. By synthesizing information from existing studies and clinical trials, I aim to elucidate key trends, challenges, and opportunities within the field. This methodology will enable me to derive actionable insights and recommendations to inform strategic decision-making for the sponsor's startup company. Through a thorough examination of peer literature and trial results, I aspire to contribute valuable knowledge that guides the sponsor in addressing unmet needs and advancing initiatives targeting rare pediatric epilepsies.

In my project focused on rare pediatric epilepsies, I plan to leverage Python for data cleaning and initial data processing, as well as Tableau for creating compelling visuals and dashboards. Python's versatility and robust libraries, such as Pandas and NumPy, will allow me to efficiently handle diverse datasets, perform data cleaning tasks, and conduct preliminary analysis. Through Python, I will address challenges like missing values, inconsistencies, and outliers, ensuring the integrity and quality of the data before further analysis. Additionally, Python's flexibility enables custom data transformations and feature engineering tailored to the specific needs of the project. Once the data is prepared, I will utilize Tableau's intuitive interface and powerful visualization capabilities to generate insightful dashboards and visuals. Tableau's rich selection of charts, graphs, and interactive features will enable me to present complex data clearly and engagingly, facilitating easy interpretation and decision-making for stakeholders.

**Preliminary Results:**

1. **Table of Year Wise Total number of Projects, Funds and Cost:**

**A table of numbers with numbers and symbols

Description automatically generated with medium confidence**

The table presents an annual overview of project counts, allocated funds, and incurred costs. Notably, the highest number of projects occurred in 2022, totaling 1,290, with an allocation of $635 million, yet only $596 million was utilized. In contrast, despite a slightly lower project count of 1,253 in 2021, funds exceeded those of 2022 by $7 million.

1. **Correlation Chart:**

**A screenshot of a graph

Description automatically generated**

The correlation matrix displayed in the heatmap visualizes the relationships between various numerical columns in the dataset. The color intensity and the annotated values represent the correlation coefficients, which range from -1 (perfect negative correlation) to 1 (perfect positive correlation). A strong positive correlation is observed between `total\_cost` and `total\_cost\_ic` (0.97), indicating that as the total cost increases, the total cost attributed to IC also tends to increase proportionally. Similarly, there is a moderate positive correlation between `direct\_cost\_ic` and both `total\_cost` (0.67) and `total\_cost\_ic` (0.68), suggesting that increases in total costs are associated with increases in direct costs to IC. The `application\_id` and `fiscal\_year` show a near-perfect correlation (0.99), likely indicating a systematic increment in application IDs over the years. Conversely, the `total\_cost\_(sub\_projects)` shows negligible correlation with most other variables, including a slightly negative correlation with `total\_cost` (-0.086) and `total\_cost\_ic` (-0.084), implying that variations in sub-project costs are relatively independent of the main project costs. This heatmap effectively highlights the key relationships among the financial aspects of the dataset, providing insights into how different cost components are interrelated.

1. **Total Cost by Fiscal Year:**

A graph of blue bars

Description automatically generated

This bar plot illustrates the expenditure on epilepsy projects from 2014 to 2023. Along the x-axis, each bar corresponds to a specific year, while the y-axis quantifies the total costs associated with these projects. Notably, the peak expenditure occurred in 2021, surpassing 6.5 billion dollars, with 2022 closely trailing at just over 6 billion dollars. A discernible trend of escalating costs emerged from 2016 onwards, culminating in the substantial investment witnessed in 2021.

1. **Institute/Center and their total cost for projects:**

A graph of blue rectangular objects

Description automatically generated with medium confidence

The above barplot shows the top 5 Administering Institutes or Centers based on their Total Cost values, with costs displayed in millions of dollars. The x-axis represents the institutes or centers, and the y-axis represents the total cost of that institute or center. NINDS stands out with a total cost significantly higher than the other institutes, about 2.5 billion dollars. This might mean that NINDS may have more funding for their projects. NIMH also has a substantial total cost but is much lower than NINDS. The other three institutes have relatively similar total costs, indicating a closer distribution of funding or expenses.

1. **No. of Projects of top institutes or centers:**

A graph of a number of blue squares

Description automatically generated

The above barplot shows the top 5 Administering Institutes or Centers based on their number of projects over the years. The x-axis represents the institute or center and the y-axis represents the number of projects. NINDS manages significantly more projects compared to the other institutes, with around 7000 projects. This indicates a high level of activity or funding allocation towards epilepsy research. NIMH follows with about 1500 projects. This suggests a strong focus on epilepsy research, but it is still substantially less active than NINDS in terms of project count. NIGMS has around 800 projects, showing a moderate level of research activity.

1. **Year Wise Activity Funding:**

**A graph of a graph showing different colored lines

Description automatically generated**

The chart provides a detailed year-wise breakdown, highlighting the disparity in total costs across different activities. The activity R01 consistently dominates the cost landscape, suggesting it may be a major focus area or a project with extensive financial requirements. The steady rise and slight recent decline in R01's costs could indicate phases of major investment followed by stabilization or reduction efforts. Other activities, while less financially intensive, show varied behavior. Activities such as U01 and UM1 exhibit noticeable peaks, suggesting specific periods of increased spending, possibly due to project milestones or new initiatives. The activities with the lowest costs remain relatively stable, implying either well-controlled budgets or smaller scope operations.

1. **Forecast:**

I have used Python to perform a time series forecast to predict the total cost for the year 2024 using historical data. It aggregates the total costs for each fiscal year using the “group by” method and sums them up. The aggregated data is then reset to a DataFrame format with each fiscal year and its corresponding total cost. Next, I created an Exponential Smoothing model to the aggregated total cost data. Exponential Smoothing is a time series forecasting method that considers trends in the data. The model is specified with an additive trend and no seasonal component. After fitting the model, the code forecasts the total cost for the year 2024 by projecting one time step ahead and prints the forecasted value. This approach helps to predict future costs based on past trends in the data.

The forecasted total cost for 2024 comes up to $698,715,288.54.

**A screenshot of a table

Description automatically generated <- Forecasted Value**

The table provides a historical record of total costs from the fiscal year 2014 to 2023 and includes a forecasted value for 2024. From 2014 to 2023, the total costs have shown a generally upward trend, with some fluctuations. Starting at $417,683,491 in 2014, the total cost dipped slightly in 2015 but then rose steadily over the years, reaching $615,231,899 in 2023. Using an Exponential Smoothing model, which accounts for trends in the data, the total cost for the fiscal year 2024 is forecasted to be approximately $698,715,289. This forecast suggests a significant increase compared to previous years, reflecting ongoing trends and potential increases in expenditures. The method used provides a statistically sound projection based on past data, which can be valuable for budgeting and financial planning.

1. **Dashboard**

A screenshot of a graph

Description automatically generated

The "Year-Wise Funding Trend" line chart depicts the total funding trend from 2014 to 2023. The graph shows a general growth in total funding, from around $400 million in 2014 to around $600 million by 2023. Particularly, funding increased significantly between 2015 and 2020, peaking around 2021, before declining somewhat in 2023. This graph successfully shows the growing trend of financing over time, with a recent drop.

The "Top 10 Institute Funding" bar chart shows the total funding received by the institutes. NINDS is the highest funded institute, receiving about $2.7 billion over the past 10 years. Following NINDS, NIMH receives the second-highest funding, which is approximately $800 million. Although it is the second highest funded institute, it is far below NINDS in terms of funds. The remaining institutes, including NIGMS, NIA, NICHD, NHGRI, NIBIB, NIDA, NIDCD, and OD, get decreasing funding, indicating a precipitous fall after the top two institutes.

The "Top 10 Institutes by Projects" bar chart shows the total number of projects managed by the top ten institutes. The NINDS has the most projects, totaling over 6,000 in 10 years. Following NINDS, NIMH manages around 1,800 projects in 10 years.

**Conclusion:**

The analysis of the dataset from 2014 to 2023 reveals several key trends and insights into epilepsy research funding and project management. Notably, the number of projects has fluctuated year by year, with a peak of 1,290 projects in 2022 and a notable funding allocation of $635 million, though only $596 million was utilized. The correlation matrix further illustrates strong positive relationships between total costs and costs attributed to intramural components, indicating that as overall project expenses increase, so do the direct costs associated with intramural activities.

The total cost analysis by fiscal year indicates a clear upward trend starting from 2016, with 2021 seeing the highest expenditure at over $6.5 billion, followed closely by 2022. This surge reflects substantial investments in epilepsy research during those years. Notably, the National Institute of Neurological Disorders and Stroke (NINDS) emerges as the top funding institute, with significantly higher total costs compared to other institutes, totaling about $2.5 billion. This suggests a substantial focus and funding commitment towards epilepsy-related projects by NINDS. It also leads with over 7,000 projects managed, demonstrating a high level of activity and funding allocation. The National Institute of Mental Health (NIMH) follows with around 1,500 projects, indicating a strong focus on epilepsy research but at a substantially lower volume compared to NINDS. The time series forecast for 2024 predicts a total cost of approximately $698 million, indicating a continued upward trajectory in funding. This forecast is based on historical trends and utilizes an Exponential Smoothing model, which accounts for trends in the data to provide a statistically sound projection.

The analysis of epilepsy research funding and project management from 2014 to 2023 offers valuable insights for Grik Therapeutics. Understanding the fluctuations in project numbers and the significant funding allocations, particularly the peak in 2022, enables Grik Therapeutics to strategically plan their research and development initiatives. The correlation between total costs and intramural costs suggests that higher project expenses often align with increased intramural activities, highlighting areas where Grik can optimize its budget. The upward trend in total cost analysis, especially the substantial investments in 2021 and 2022, indicates a growing focus on epilepsy research, which Grik can leverage for potential funding opportunities. Notably, the National Institute of Neurological Disorders and Stroke (NINDS) emerges as a major funder, suggesting that aligning with NINDS's priorities could be beneficial. Additionally, with NINDS managing over 7,000 projects, there are ample opportunities for collaboration and resource sharing. The forecasted increase in total costs for 2024 further underscores the expanding landscape of epilepsy research funding, positioning Grik Therapeutics to secure necessary resources for its innovative projects.

s

**References:**

1. Ramuthi, D. (2024, April 19). *How to Write a Research Proposal: A Step-by-Step - Venngage*. Venngage. <https://venngage.com/blog/how-to-write-a-research-proposal/#:~:text=Your%20proposal%20should%20also%20outline%20the%20research,research%20schedule%2C%20and%20a%20budget%20if%20applicable>.
2. Privitera, M. (2006, January 1). *Large clinical trials in epilepsy: Funding by the NIH versus pharmaceutical industry*. Epilepsy Research. <https://doi.org/10.1016/j.eplepsyres.2005.09.020>
3. Zurdo-Tabernero, M., Canal-Alonso, N., De La Prieta, F., González, S. R., Prieto, J., & Corchado, J. M. (2023, December 15). *An overview of machine learning and deep learning techniques for predicting epileptic seizures*. Berichte Aus Der Medizinischen Informatik Und Bioinformatik/Journal of Integrative Bioinformatics. <https://doi.org/10.1515/jib-2023-0002>
4. Kumari, K. (2024, April 30). Time Series Forecasting Using Python. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2022/06/time-series-forecasting-using-python/>