Capstone Project Data 205: Data Story

Title: Datamining for OSEL Regulatory Science Programmatic Planning

In the ever-evolving landscape of medical technology, the Office of Science and Engineering Laboratories (OSEL) embarked on a critical journey to pinpoint the gaps in regulatory science, mainly focusing on the cardiovascular program, which aimed to address a crucial need in the realm of medical device regulation - keeping pace with rapid scientific and technological advancements. Also, to seek potential research collaborators.

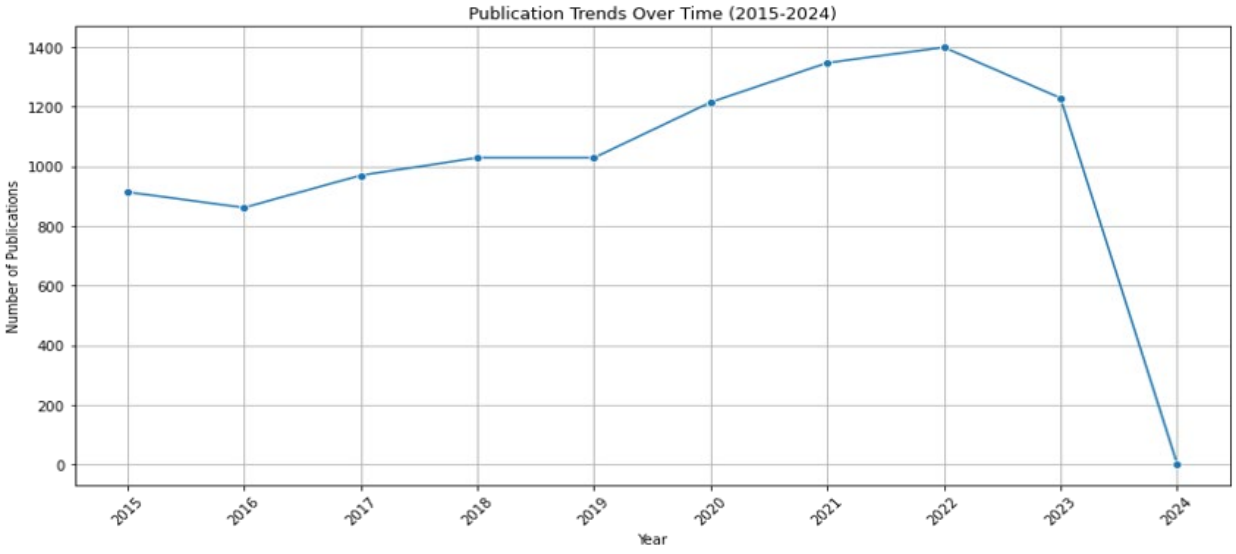
The concept of a regulatory science gap is central to understanding the impetus behind OSEL's project. It refers to the areas within regulatory science where there is a lack of knowledge, understanding, methods, or tools. For example, in the cardiovascular field, this manifested as several challenges. These included outdated techniques for assessing blood damage preclinically, an over-reliance on clinical data for evaluating hemodynamics, concerns about the durability of cardiovascular implants, the unpredictability of cardiac medical devices' success, and the absence of human physiology-based testing methods.

However, addressing these challenges required an innovative approach, and OSEL rose to the occasion by employing a unique data mining strategy. The traditional avenues for data collection proved inadequate, necessitating a shift towards more advanced techniques. Python programming and the BeautifulSoup library developed a web scraping algorithm. This algorithm was designed to extract pertinent data from PubMed, a renowned medical research database. The search was meticulously planned, focusing on terms like 'Hemolysis' and 'blood damage' in conjunction with specific medical devices and conditions such as 'Heart Valves,' 'Ventricular Assist Devices (VADs),' 'Dialyzers,' 'Endovascular Stents,' and 'Inferior Vena Cava.' This

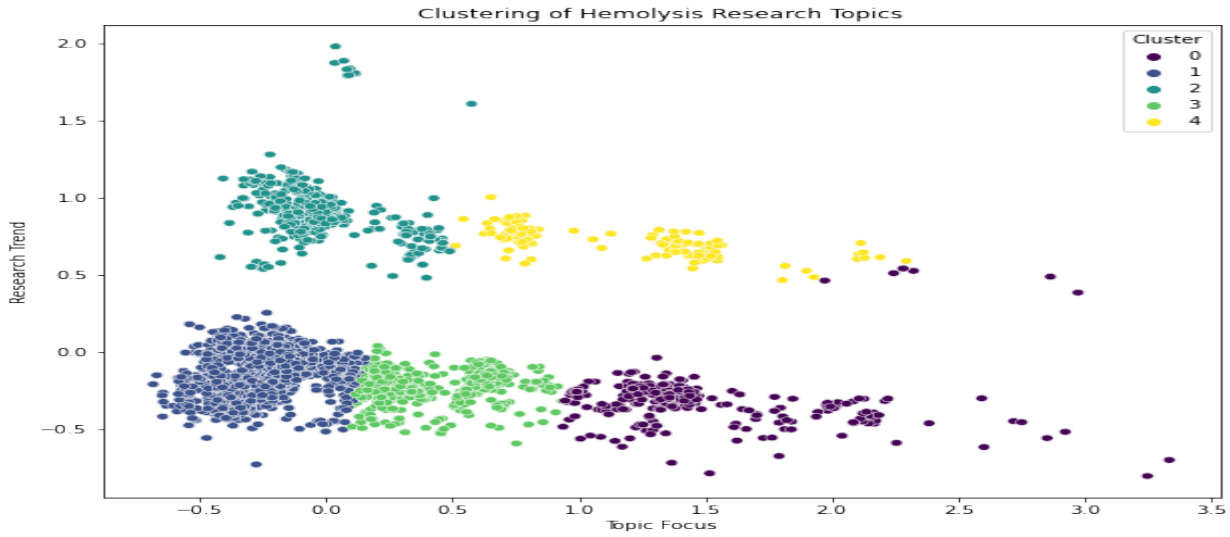
systematic approach yielded a spectrum of findings, ranging from thousands of papers on general topics to singular results for more focused queries. This strategy facilitated efficient data collection and set a new benchmark for research methodologies in related medical fields.

The datasets from this extensive data mining exercise were rich in information. They included bibliographic details of scientific publications within the field of biomedicine, uniformly featuring columns such as 'Title,' 'Authors,' 'Journal,' 'PMID,' 'Publication Year,' and 'Affiliations.' This consistency was crucial for ease of comparison across various research topics. Undertook rigorous data cleaning and preparation processes, ensuring the removal of duplicates, checking for missing values, and standardizing text data. A step further was taken by merging these datasets to conduct a comprehensive analysis comparing various medical devices. This helped identify gaps in the research publication related to specific devices and classify the medical devices under study.

The statistical analysis of this data was particularly enlightening. Visualizations, including line graphs, provided clear insights into publication trends. A consistent increase in articles from 2015 to 2022 indicated a growing research interest in this field. However, a slight decrease was noted in 2023, and the data for 2024 was incomplete.

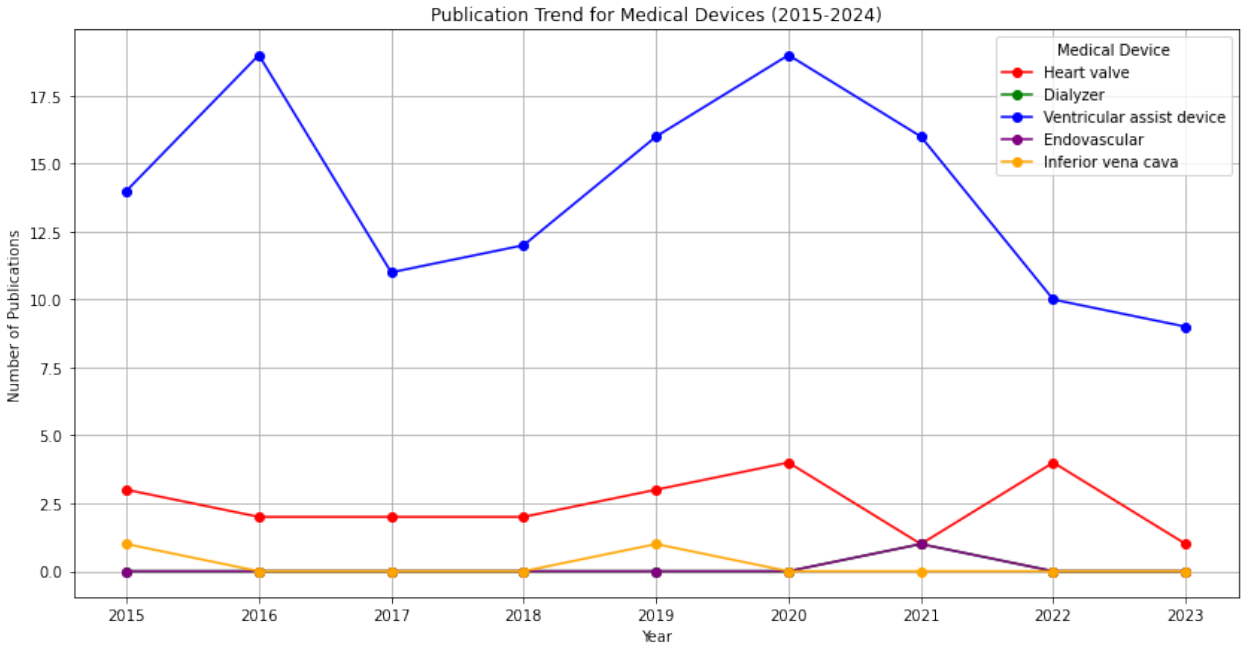


Moreover, a Chi-Square test was conducted to explore any potential seasonal patterns in publication dates, which found no significant association, thereby debunking any hypothesis of seasonal bias in research activities. Also, a clustering analysis was conducted by applying the KMeans algorithm to categorize scholarly article titles; we discern five distinct clusters that emerge, each representing a nuanced avenue of study within the field. The visualization eloquently captures the essence of these groupings, with the proximity and color coding of points revealing the thematic relationships between papers. Notably, there's a vibrant tapestry of topics ranging from the physiological impact of medical interventions on blood cells to the intricate dynamics of blood storage and transfusion. This methodical partitioning not only highlights the prevailing focal points of contemporary research but also subtly suggests potential gaps and frontiers ripe for exploration, painting a multifaceted picture of the scientific dialogue unfolding in this vital area of healthcare.



Additionally, in our thorough search to uncover any potential gaps in regulatory science regarding medical devices, we delved into a detailed analysis focused on the medical device sector. Our objective was to pinpoint areas where further investigation and regulation might be needed,

particularly for devices like heart valves and the inferior vena cava. To guide us through this complex landscape, we crafted a line graph that maps out the level of research activity for different types of devices, serving as a navigational tool to highlight where more attention and resources might be necessary.



The graph delineates an apparent disparity in research focus. Ventricular assist devices and heart valves, represented by the soaring lines, indicate a robust and sustained interest, underscoring their critical role in addressing cardiac ailments. This concentrated research effort proactively responds to the demand for advanced cardiac support technologies.

Conversely, dialyzers and technologies about the inferior vena cava, illustrated by the more modest undulations on the graph, exhibit comparatively less research activity. This suggests either a satisfactory understanding of these devices or potential oversight, signaling an area where regulatory science may need to intensify its scrutiny.

The narrative turns curious with the stark absence of data points for endovascular grafts. This void in the graph does not merely represent a lack of data; it symbolizes the silent spaces in our collective knowledge, challenging us to delve deeper. It prompts the question of whether this gap is due to a nascent stage of development in this domain or a blind spot in our current research focus.

Thus, the line graph does not merely chart the present; it beckons us toward the future, urging us to marshal our resources to fill these gaps, ask the difficult questions, and commit to a path of discovery that ensures regulatory science keeps pace with the rapid evolution of medical devices.

In sum, the final output of this project was not just a set of data but a comprehensive suite of tools and insights. It included visualizations highlighting publication trends and research gaps and a list of potential collaborators in the USA. This aspect was precious as it opened doors for future partnerships and research endeavors.

	USA Affiliations
0	Abbott Laboratories, Chicago, IL
1	BioCirc Research Laboratory, School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, Pennsylvania
2	Biomedical Acoustics Research Laboratory, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816
3	Cardiac Biomechanics Laboratory, Department of Mechanical and Materials Engineering, University of Denver, 2390 S. York St. #200, Denver, CO, 80210
4	Cardiac Ultrasound Laboratory, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Yawkey 5B, Boston, MA, 02114
5	Cardiothoracic Surgery, Crozer-Chester Medical Center, Upland
6	Cardiovascular Surgery Research Laboratories, Texas Heart Institute, Houston, TX
7	Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland
8	College of Medicine, St. Christopher's Hospital for Children, Drexel University, Philadelphia, Pennsylvania
9	College of Medicine, University of Central Florida, 6850 Lake Nona Blvd, Orlando, FL 32827
10	Computational Fluid Dynamics Group, Center for Advanced Vehicular Systems, Mississippi State University, Starkville, Mississippi
11	Department of Aerospace Engineering, Auburn University, Auburn, AL
12	Department of Biochemistry, University of Mississippi Medical Center, Jackson, MS
13	Department of Bioengineering, Speed School of Engineering, University of Louisville, Louisville, KY, 40292
14	Department of Biology, Tougaloo College, Tougaloo, MS
15	Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania
16	Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA
17	Department of Biomedical Engineering, Kate Gleason College of Engineering, Rochester Institute of Technology, Rochester, New York
18	Department of Biomedical Engineering, Michigan Technological University, Houghton, MI
19	Department of Biomedical Engineering, The Ohio State University, 473W 12th Ave., Columbus, OH, 43210

To summarize, OSEL's data mining project was not just a triumph in achieving its objectives. It represented a significant step forward in medical device regulatory science. By effectively identifying gaps in research and potential collaborators, the project fulfilled its initial goals and laid the groundwork for future advancements in this sector, and the value of it extends beyond the scientific community, offering tangible benefits to the general population. It enhances medical devices' safety and efficacy, directly impacting public health and well-being. Also, the insights gained from this project can guide patients in making more informed decisions regarding medical treatments and contribute to advancing medical research and innovation. In essence, this project stands as a testament to the power of data mining in bridging the gaps between technology, regulation, and public health.