Analysis and Trend Evaluation of Medical Device Recalls

Hunter Belous, Jared Fox, Asad Kamal
Milestone 1 - MADS Degree Candidates

Overview

Manufacturers are not always able to catch defects or adverse effects during the lifecycle of a medical device, from raw material sourcing to patient use. Many issues, such as design flaws, manufacturing errors, or labeling inconsistencies, can go unnoticed until devices are already being utilized, leading to recalls and safety alerts. Using data from the ICIJ's International Medical Devices Database (IMDD), which contains over 120,000 records of recalls and safety notices worldwide, an analysis of these common causes reveals patterns of frequent or severe recalls across regions, device categories, and manufacturers. This transparency helps drive improvements in both regulatory oversight and the development of safer, more reliable medical devices.

Project Objective

The purpose of this study is to analyze medical device recall trends and identify the most common causes, such as design flaws, manufacturing errors, and labeling issues. By comparing recall rates across different regions, device categories, and manufacturers, the study aims to uncover patterns that contribute to frequent or severe recalls. This research intends to highlight areas where improvements in device development and oversight can reduce patient risks and enhance overall medical device safety throughout their lifecycle.

Applicable Stakeholders

The findings from this data are intended to inform key stakeholders, including regulatory bodies, OEMs, contract manufacturers, healthcare providers, and patients directly affected by the recalls. The visualizations aim to present clear, concise insights into which entities are responsible for various recalls, helping stakeholders better understand their role and impact within the industry.

Johnson & Johnson

stryker

ZIMMER BIOMET

smith&nephew

invima
Instituto Nacional de Vigilancia de Medicamentos y Alimentos.

FDA
The Food and Drug Administration

Data Sources

File Name	Description	Size	URL
1991-2022 IMDD events-1681209680.csv	The Events dataset captures detailed information about specific actions or incidents related to medical devices. It includes columns that describe the nature of the event (action, action_classification, action_level, action_summary), along with several key dates that track the event's lifecycle (create_date, date_initiated_by_firm, date_posted, date_terminated, date_updated). This dataset also includes details about the cause of the event (determined_cause), the reason for the event (reason), and relevant documentation or links to authorities (documents, authorities_link). Additionally, geographical information (country), event status (status), type (type), and target audience (target_audience) provide further context. Unique identifiers such as id, uid, uid_hash, and device_id allow for precise tracking and cross-referencing with other datasets.	68.7 MB	https://medicaldevices.icij.org/p/about
IMDD devices-1681209661.csv	The Devices dataset provides comprehensive details about the medical devices themselves. Key attributes include the device name (name), description (description), and classification (classification). The dataset also includes the risk level associated with the device (risk_class), whether the device is implanted (implanted), and the quantity in commerce (quantity_in_commerce). Geographic distribution is captured through the distributed_to and country columns, and the manufacturer_id column links each device to its corresponding manufacturer in the Manufacturers dataset. The dataset also tracks the creation and modification dates for each device record (created_at, updated_at).	149.2 MB	https://medicaldevices.icij.org/p/about
IMDD manufacturers-1681209657.csv	The Manufacturers dataset details information about the companies responsible for producing medical devices. Each manufacturer is identified by a unique id, with additional details like the company name (name), address (address), and parent company (parent_company). The representative column provides the contact information for a key person at the company. This dataset also includes timestamps (created_at, updated_at) to track when records were created or last updated. The manufacturer_id column links this dataset to the Devices dataset, enabling a connection between devices and their manufacturers.	9.5 MB	https://medicaldevices.icij.org/p/about

Initial Dataset Manipulation

Manipulation Techniques with Python in Jupyter Notebooks

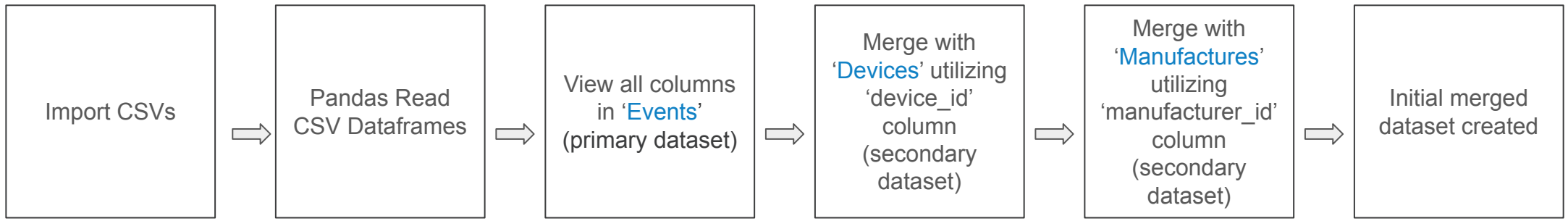
To effectively preprocess the data from the IMDD dataset, Python was used within Jupyter notebooks. This approach allowed for clear and concise documentation of the steps taken to reach the conclusions presented in this study. As detailed in the process flow below, the pandas library was utilized to load publicly available CSV files and convert them into pandas DataFrames, enabling further data manipulation and analysis. This ensured the data was properly processed for subsequent exploration.

Approach

Instead of filtering each dataset individually before merging, the IMDD, provided by the International Consortium of Investigative Journalists (ICIJ), compiles data from various global regulatory sources to create a unified dataset. The ICIJ is a nonprofit organization known for investigative reporting on cross-border issues, including medical device safety. The information was supplied in three separate files for user manipulation. Using device identification codes and manufacturer names, events in the primary dataset were successfully matched with corresponding records in the secondary dataset. After merging the three datasets, cleaning and filtering methods were applied to prepare the data for analysis.

Data Merging and Cleaning (1/2)

The merging process was carried out using the device_id (from the device dataframe) and manufacturer_id (from the manufacturer dataframe) columns to link the three datasets in total: **Events**, **Devices**, and **Manufacturers**. This step allowed for the consolidation of related records across different datasets, creating an initial dataframe for each event to be associated with the correct device, manufacturer, and all applicable information.



Finalized Dataset Manipulation

Data Merging and Cleaning (2/2)

1. Duplicates can arise from data merging processes or data entry error. Duplicate records were removed with `final_df.drop_duplicates(inplace=True)` to ensure that each data point is unique, preventing skewed analysis results.
2. Data columns were converted to datetime format with `final_df['date'] = pd.to_datetime(final_df['date'], errors='coerce')` which allows for accurate date-based analysis and computations, such as time series analysis and trend evaluation. The `errors='coerce'` parameter handles invalid date formats by converting them to NaT (Not a Time).
3. Column names were standardized with `final_df.columns.str.lower().str.replace(' ', '_')` to achieve a consistent format of lowercase characters with underscores instead of spaces. This ensures consistency and ease of access, making the DataFrame easier to work with in a coding environment.

Handling Missing Values

1. Missing values were initially identified using `final_df.isnull().sum()` to understand the extent and distribution of them and to decide on an optimal and balanced approach. Dropping all records with missing values would lead to significant data loss, which might undermine the analysis.
2. Ensuring that essential columns (such as `device_id, manufacturer_id, reason`) have no missing values maintains the integrity of the analysis. These columns are critical for identifying and understanding medical device recall events. Records with missing values from these essential columns were dropped from the final DataFrame with `final_df.dropna(subset=essential_columns, inplace=True)`
3. Missing values from other important categorical columns such as `classification, determined_cause, description` were imputed with the mode to ensure that these variables remain useful for analysis without introducing bias. It was decided to utilize the mode as it maintains the distribution's integrity. A for loop was implemented to iterate over each column in the `important_categorical_cols` list. For each column, the mode was calculated using `final_df[col].mode()`.
4. Columns with a large proportion of missing values such as `data_notes, date_posted, date_updated`, and others were dropped, allowing the analysis to be focused on the most relevant data while simplifying the dataset. Columns with significant missing data add noise and complexity without any benefit.
5. The final dataset was validated using `final_df.isnull().sum()` to ensure that all missing values were addressed appropriately. Initially, the merged DataFrame had a shape of (124,969, 55). After applying the data cleaning and manipulation steps, the shape of the final DataFrame became (9,067, 42), indicating that we have retained a substantial and statistically significant amount of data for making informed decisions and drawing conclusions from the analysis.

Frequency Analysis of Medical Device Recalls

Frequency Analysis of Medical Device Recalls

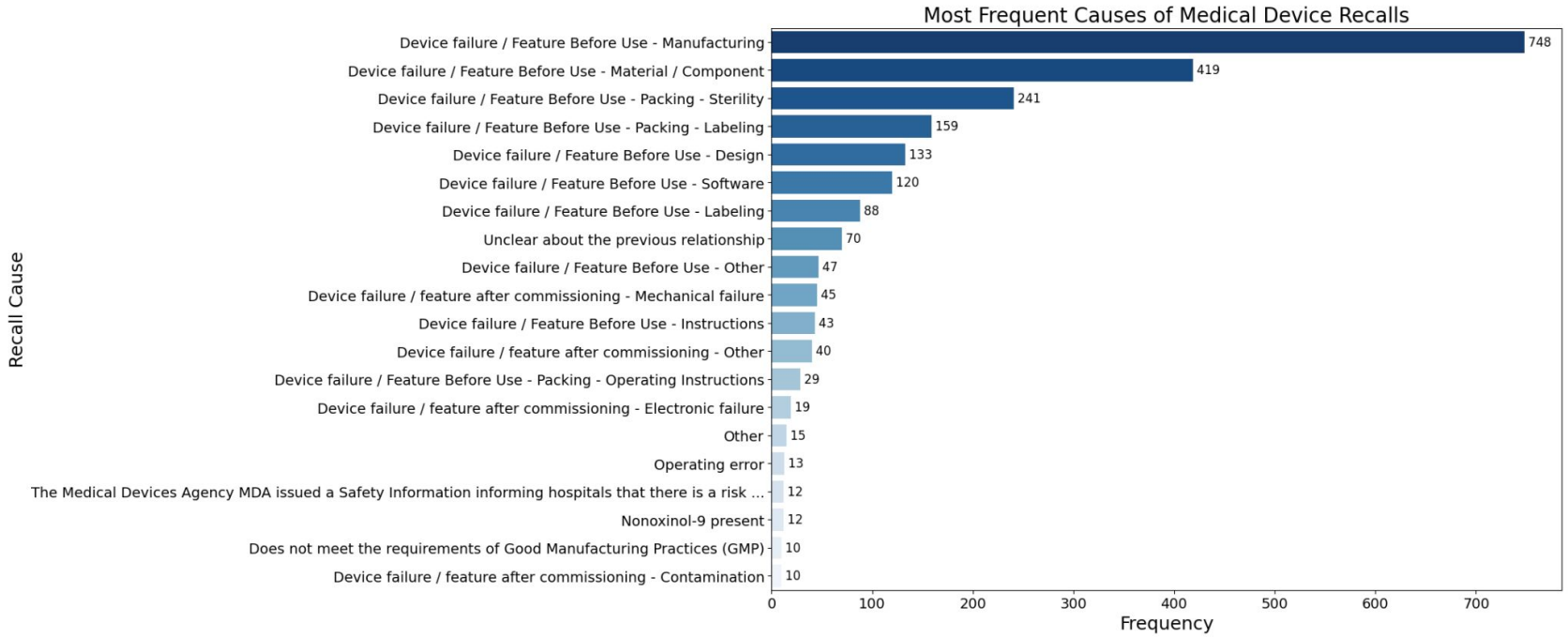
A frequency analysis was conducted to explore and understand the common causes of medical device recalls. By identifying the most frequent recall causes, we aim to gain insights into the predominant issues necessitating product recalls, and to inform strategies for improving device safety and compliance. For clarity, the analysis was focused on the top 20 most frequent recall causes.

Key Findings

The most common cause of medical device recalls is related to manufacturing defects that lead to device failure before use. Manufacturing defects can include inadequate production processes, faulty assembly, or issues with the quality control of materials. This high frequency indicates a critical need for stringent manufacturing quality assurance and control processes.

Significantly, the second most frequent cause pertains to failures in materials and components, highlighting the importance of ensuring the integrity of all raw materials and components used in manufacturing.

Additional significant recall causes are mostly related to device failure before use - packing sterility, labeling, design, software, etc. Other general failures, operating instructions issues, and specific labeling failures show between 10 to 45 occurrences.



Recall Cause	Frequency	Percentage
Device failure before use	1940	91%
Device failure after use	124	5%
Other	82	4%

Insights from Device Failure Analysis

Device Failure After Use

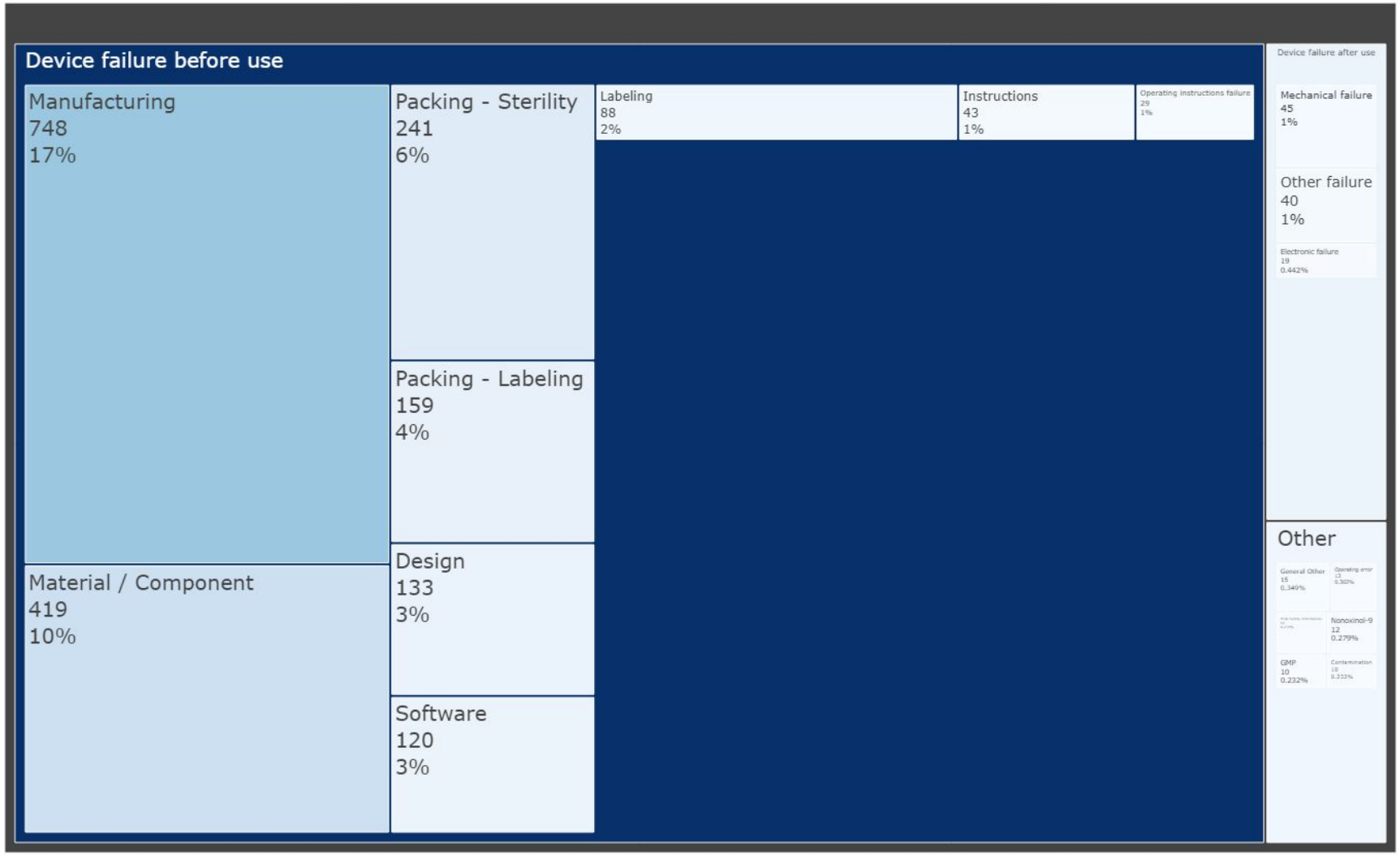
From the hierarchical representation of the treemap, it is evident that within the category of device failure before use, the largest subcategories include manufacturing defects, material and component issues, and sterility problems. Another significant section of the treemap is dedicated to device failure after use, highlighting issues detected post-usage of the device.

Mechanical failures, comprising 45 occurrences, indicate potential weaknesses in the durability and operational reliability of the devices. The 'other' category encompasses a variety of recall reasons that do not fit into the primary categories. This includes user-related errors, such as operating errors (13 occurrences), which suggest a need for improved user training or clearer instructions.

Additionally, it is important to recognize that a medical device recall may entail various corrective actions, such as repairing or adjusting the device, re-labeling, or other necessary interventions. The FDA (Food & Drug Administration) classifies recalls based on the risk they pose:

- **Class I**: Reasonable chance of causing serious health issues or death.
- **Class II**: May cause temporary or reversible health problems or slight chance of serious health issues or death.
- **Class III**: Not likely to cause any health problems or injury.

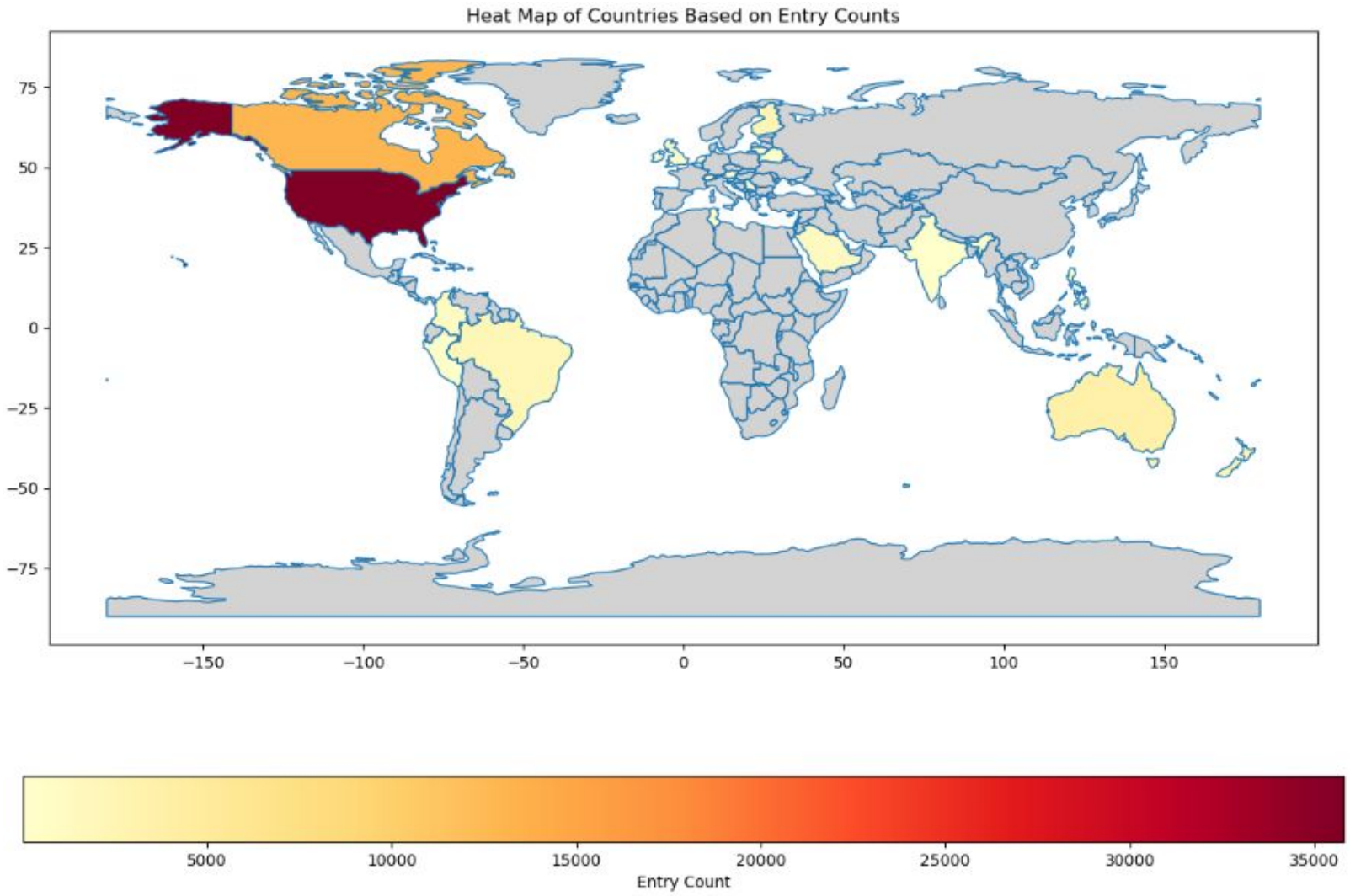
Medical Device Recall Causes Treemap



Medical Device Recalls by Region

World Map of Recalls

To get a better understanding of correlations between recall occurrence and geographical location, a world heat map was created. This visualization shows that a majority of the datasets recalls occurs within the United States and Canada. The largest 3 areas were the United States, Canada, and Australia, with 35799, 12901, and 3558 entries respectively. This created a glaring issue. With the two major outliers of the United States and Canada, the heat map really couldn't provide much information besides the countries involved in the dataset and the two major countries being much more of the dataset than the others. To resolve this issue, another model was created with the outliers removed.



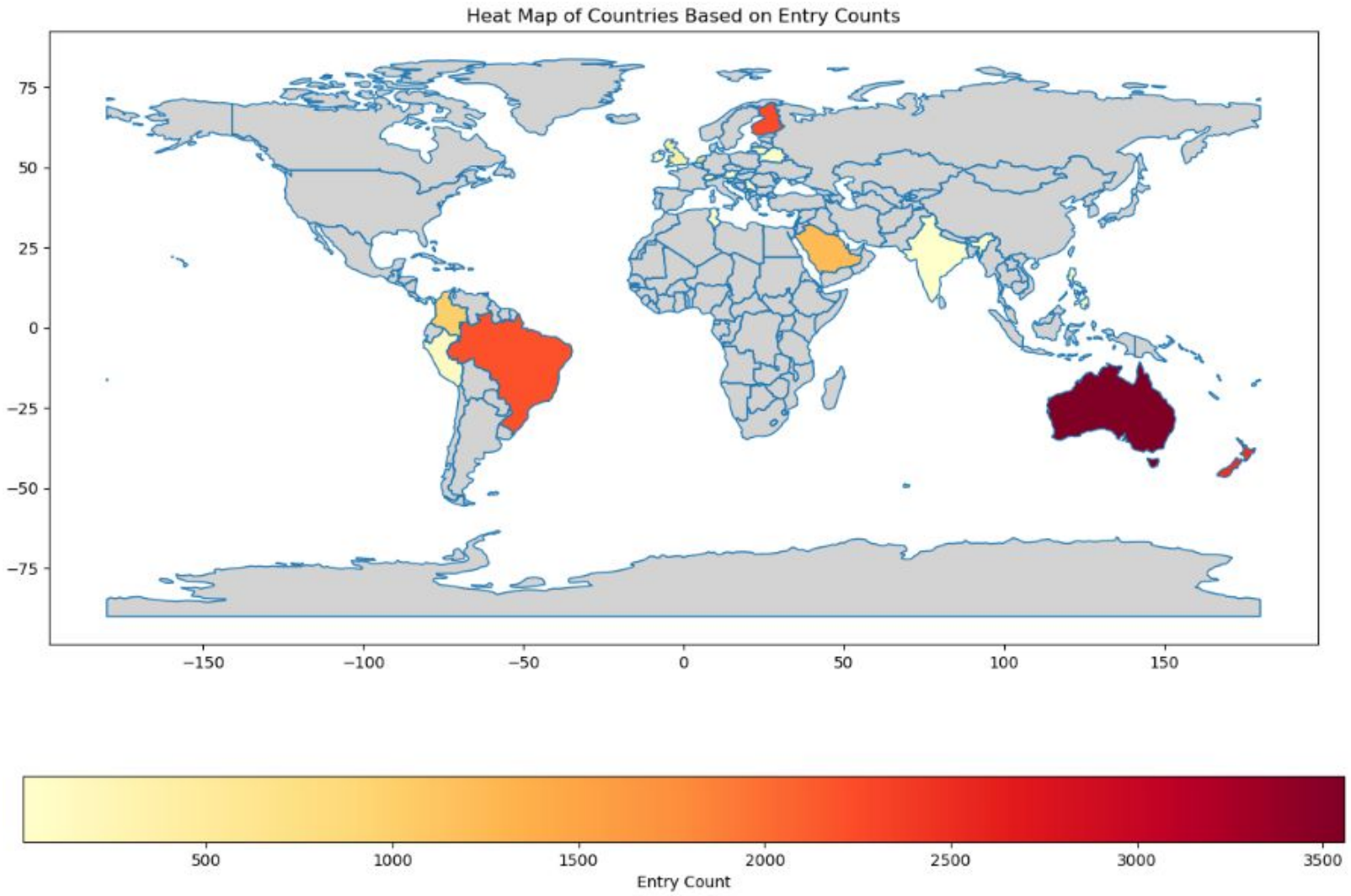
Medical Device Recalls by Region

World Map of Recalls without Outliers

With the outliers removed, the heatmap created a much clearer picture of what countries were involved in the recall dataset and how they compared to each other. The model shows that many of the countries with high volumes of recalls are mixed between wealthy countries and less prosperous countries. It also shows that geographical region does not appear to have an affect on recall rate.

Limitations

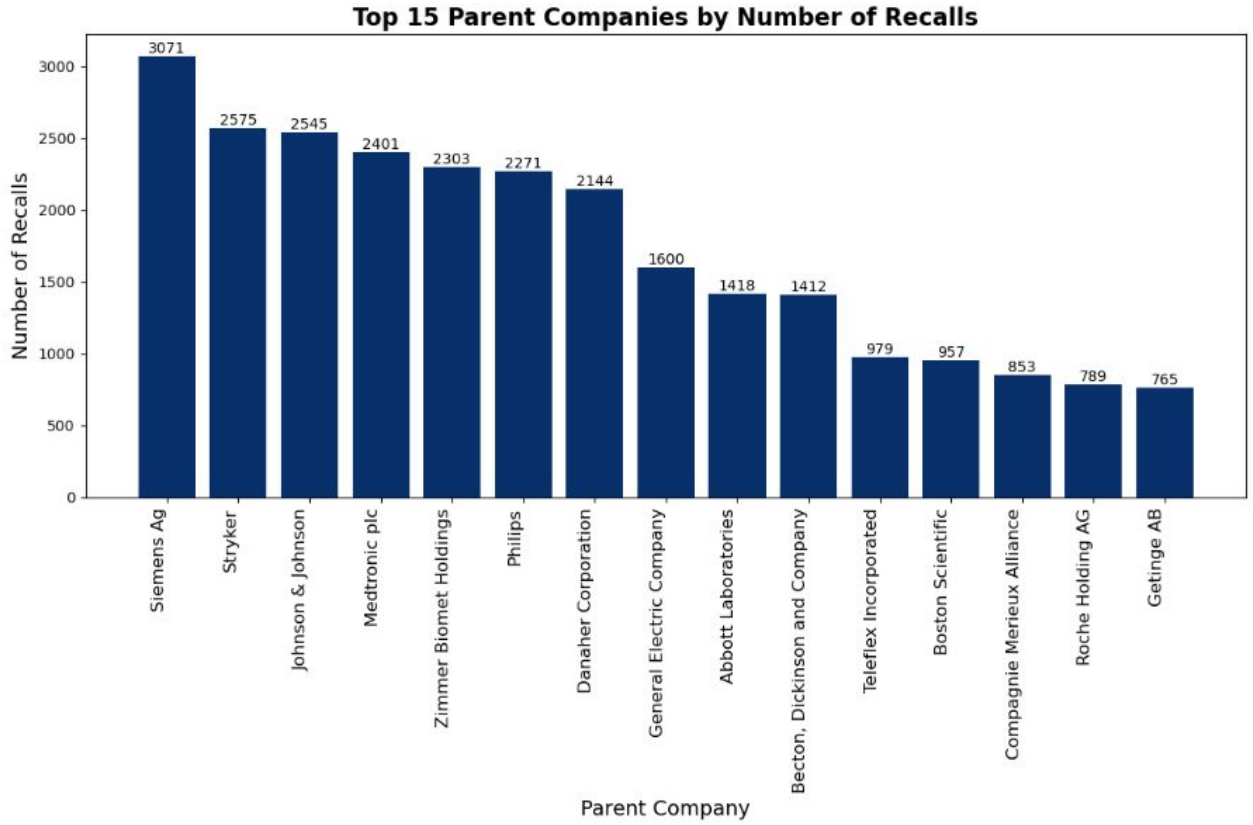
There are many limitations with this model. There are countries that most likely do not publish data. Another major limitation is the limited outside knowledge on which country has the most manufacturing sites or which sites receive medical recalls.

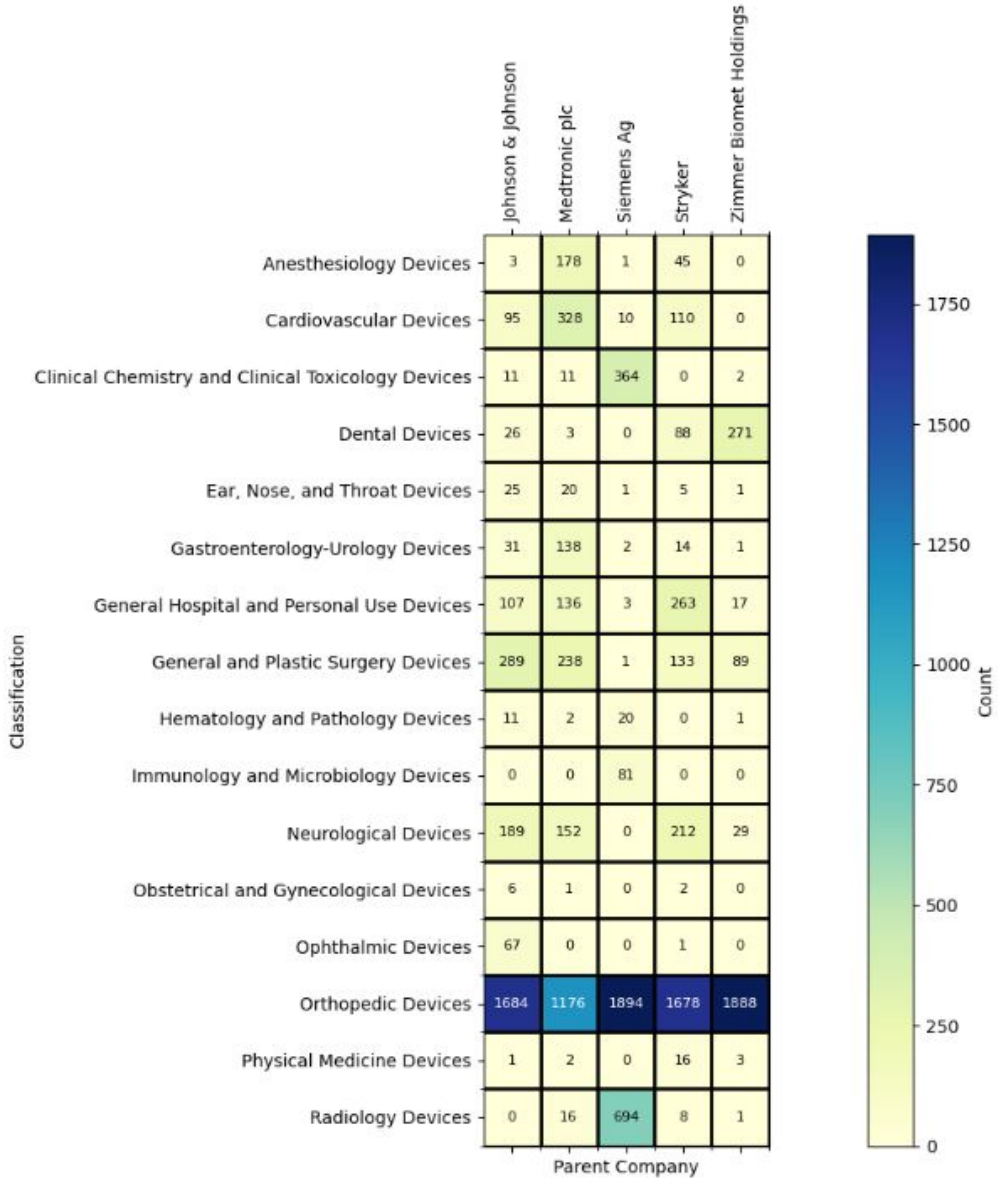


Medical Device Recalls by Parent Companies

Medical Device Recalls by Company

After grouping the data by frequency of recalls by company, the following bar graph displays the top 15 parent companies by total number of medical device recalls, with Siemens Ag leading at 3,071 recalls, followed by Stryker and Johnson & Johnson. Orthopedic Devices appear to be key contributors to the high recall numbers for these companies. Companies like Medtronic and Zimmer Biomet Holdings also report significant recall volumes across various classifications. Moving to the contingency table, it shows how specific device categories, such as Orthopedic devices, have high recall concentrations for certain companies. The chi-square test further confirms the significance of these relationships, with a chi-square statistic of 7645.94 and a p-value of 0.0, indicating a very strong association. This statistical significance suggests that certain companies are more likely to have recalls in particular device categories, pointing to areas that may require further investigation or process improvements.





Summary

Summary of Findings and Insights on Medical Device Recalls

Our comprehensive analysis and trend evaluation of medical device recalls have revealed significant insights into the predominant issues and potential areas for improvement. Here are the key findings and insights drawn from the various analytical steps:

Objective and Overview:

- **Purpose:** Analyze medical device recall trends to identify common causes, such as design flaws, manufacturing errors, and labeling issues, and compare recall rates across different regions, device categories, and manufacturers.
- **Data Source:** The ICIJ's International Medical Devices Database (IMDD) provided data on over 120,000 recalls and safety notices worldwide.

Data Cleaning and Manipulation:

- Adopted strategic approaches to handle missing values, ensure data integrity, and maintain a balanced dataset.
- Removed duplicates, converted date columns to datetime format, and standardized column names.
- Dropped columns with extensive missing values while imputing critical missing data with the mode for consistency.
- Resulted in a clean, validated dataset of 9,067 records, ensuring statistical significance and robustness for analysis.

Frequency Analysis:

- **Top Recall Cause:** Manufacturing defects leading to device failure before use were identified as the most common recall cause (748 occurrences).
- **Other Frequent Causes:** Failures related to materials/components, packaging sterility, labeling issues, and design flaws were also significant.
- Highlighted the critical need for stringent manufacturing quality assurance and control processes.

Geographical and Manufacturer Analysis:

- **Manufacturer Trends:** Top companies with highest recalls included Siemens AG (3,071 recalls), Stryker, and Johnson
- **Regional Insights:** Initial heat map showed major recalls in the United States (35,799) and Canada (12,901), with significant entries also in Australia (3,558). A subsequent map without these outliers provided a clearer global recall pattern.

Statement of Work/References

Hunter Belous

Hunter Belous was responsible for all visualizations in the *Medical Device Recalls by Parent Company* section, as well as the introduction slide, data source slides, and initial manipulation slide. His analysis uncovered key trends among parent companies and their related recalls.

Jared Fox

Jared Fox was responsible for all visualizations in the Medical Device Recalls by Region section, as well as the summarization slide of the report. He also played an integral role in refining the dataset to enhance each visualization. His insights provided a clearer understanding of the global distribution of recalls.

Asad Kamal

Asad Kamal was responsible for all visualizations within the Frequency Analysis and Insights from Device Failure Analysis sections. He is the author of the dataset merging and cleaning notebook. His analysis revealed the distribution of medical device recalls as well as implications for next steps from the analysis.

References

International Consortium of Investigative Journalists. (2024). *Devices.* Retrieved from https://medicaldevices.icij.org.

International Consortium of Investigative Journalists. (2024). *Events.* Retrieved from https://medicaldevices.icij.org.

International Consortium of Investigative Journalists. (2024). *Manufacturers.* Retrieved from https://medicaldevices.icij.org.

U.S. Food and Drug Administration. (n.d.). Medical device recalls. Retrieved from https://www.fda.gov/medical-devices/medical-device-safety/medical-device-recalls.