

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

## **Data and their Sources**

*1st data source:*

<https://www.kaggle.com/datasets/prasad22/ca-independent-medical-review>

California permits reviews of an insurance company's denial of services as per an individual's health insurance plan. This is managed by the California Department of Managed Health Care (DMHC) via an Independent Medical Review (IMR) board. An IMR is a review of a denied, delayed, or modified health care service that the health plan has determined to be unnecessary. If the IMR decides in an enrollee's favor, the health plan must authorize the service or treatment requested. This data is from the California Department of Managed Health Care (DMHC). It contains 19,245 decisions from January 1, 2001 to 2016.

Definition of the structured data:

Category	Description
Reference ID	Unique ID of the review
Report Year	The year in which the review was performed
Diagnosis Category	The diagnosis of the patient related to the type of physician
Diagnosis Sub-Category	The detailed diagnosis within the diagnosis category
Treatment Category	The treatment sought for the patient
Treatment Sub-Category	The detailed treatment within the specific Treatment Category
Determination	The ruling of the IMR board (either overturned or upheld)
Type	The type of treatment (urgent care, medical necessity, experimental)
Age Range	Various age groupings from 0-10 through 65+
Patient Gender	Male or Female

## **Unstructured data:**

Included in the categorized data above, is a written explanation for each of the 19,245 decisions referred to as "Findings". These explanations are free form text consisting of ~300 words each (~6MM words across the entire data set). This is the equivalent of about 18 novels at 1000 pages each.

## **Sample of text data within the "Findings" column of each review:**

K
Findings
Nature of Statutory Criteria/Case Summary: An enrollee has requested Harvoni for treatment of his hepatitis C virus genotype 1a. Findings: The physician reviewer found that according to the most recent joint guidelines issued by the American Association for the Study of Liver Diseases (AASLD) and the Infectious Diseases Society of America (IDSA), all patients with chronic hepatitis C should be treated except those with limited life expectancy due to non-liver-related conditions. This applies regardless of fibrosis stage or viral load, and advanced fibrosis is not required for treatment. Per guidelines, treatment-naïve genotype 1 patients should be treated with Harvoni. Patients with viral load of less than 6,000,000 IU/mL can be treated for eight weeks. These guideline recommendations are based on multiple randomized clinical trials (Kowdley, et al; Afdhal, et al). Moreover, the AASLD and IDSA guidelines cite data showing reduced survival with delayed treatment (Jezequel, et al). For these reasons, the requested treatment with Harvoni is supported as medically necessary for the treatment of this patient's hepatitis C infection. Final Result: The reviewer determined that the requested medication is medically necessary for treatment of the patient's medical condition. Therefore, the Health Plan's denial should be overturned. Credentials/Qualifications: The reviewer is board certified in internal medicine with sub-specialty certification in gastroenterology and is actively practicing. The reviewer is an expert in the treatment of the enrollee's medical condition and knowledgeable about the proposed treatment through recent or current actual clinical experience treating those with the same or a similar medical condition.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

*2nd data source:*

<https://data.cdc.gov/api/views/g4ie-h725/rows.json?accessType=DOWNLOAD>

The first dataset utilized in this analysis originates from the Data.gov data catalog, specifically sourced from the CDC's Division of Population Health. This division has established a comprehensive collection of 124 indicators, which have been carefully developed through consensus. These indicators enable consistent definition, collection, and reporting of chronic disease data, thus supporting public health practices across various states, territories, and large metropolitan areas. In addition to offering access to state-specific indicator data, the CDI website also serves as a valuable gateway, providing further information and additional data resources.

### **Data Exploration and Cleaning**

The categorized data was well-organized and presented in a tabular format. However, there was a high number of NaN values. To address this, we calculated the total NaNs per attribute and their percentages relative to the entire dataset. We also utilized the information from the Findings column to fill in some of the missing values for Age Range and Patient Gender. Updated the Age Range category to use a dash instead of an underscore for the 11-20 bin, in order to make it consistent with the other bins.

*Second dataset:*

This data was imported from a JSON formatted file. A custom function was developed to read the JSON data and display its hierarchical structure. This helped identify the necessary name-value pairs for analysis. Once the relevant data was extracted from the JSON file, it was loaded into a pandas dataframe.

Subsequent exploratory analysis was conducted to identify any issues such as missing values (NaN), infinite values (inf), or other discrepancies in the dataset. Columns with a significant number of missing values and observations without any measurements were removed from further analysis. To enhance the dataset, feature engineering techniques were applied. Specifically, two new columns, "year\_bins" and "us\_regions," were created. Since some measurements covered two, three, or five-year periods, bins were introduced to indicate this temporal aspect. Additionally, a "region" column was added to categorize states based on their respective regions.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

## **Ryan's Analysis**

### **Comparison Questions #1**

Are there any differences in the IMR determination outcomes for different types of medical services?

- Unit of analysis: Medical Service Type and Determination
- Comparison values:
  - Raw counts of cases overturned compared to upheld based on medical service type.
  - Percentage of cases overturned compared to upheld based on medical service type.
- Computations:
  - Bar chart
  - Contingency table
  - Chi-square test of independence to determine if there's an association between IMR determination and whether the case is overturned or upheld.

### **Brief Description of the Program**

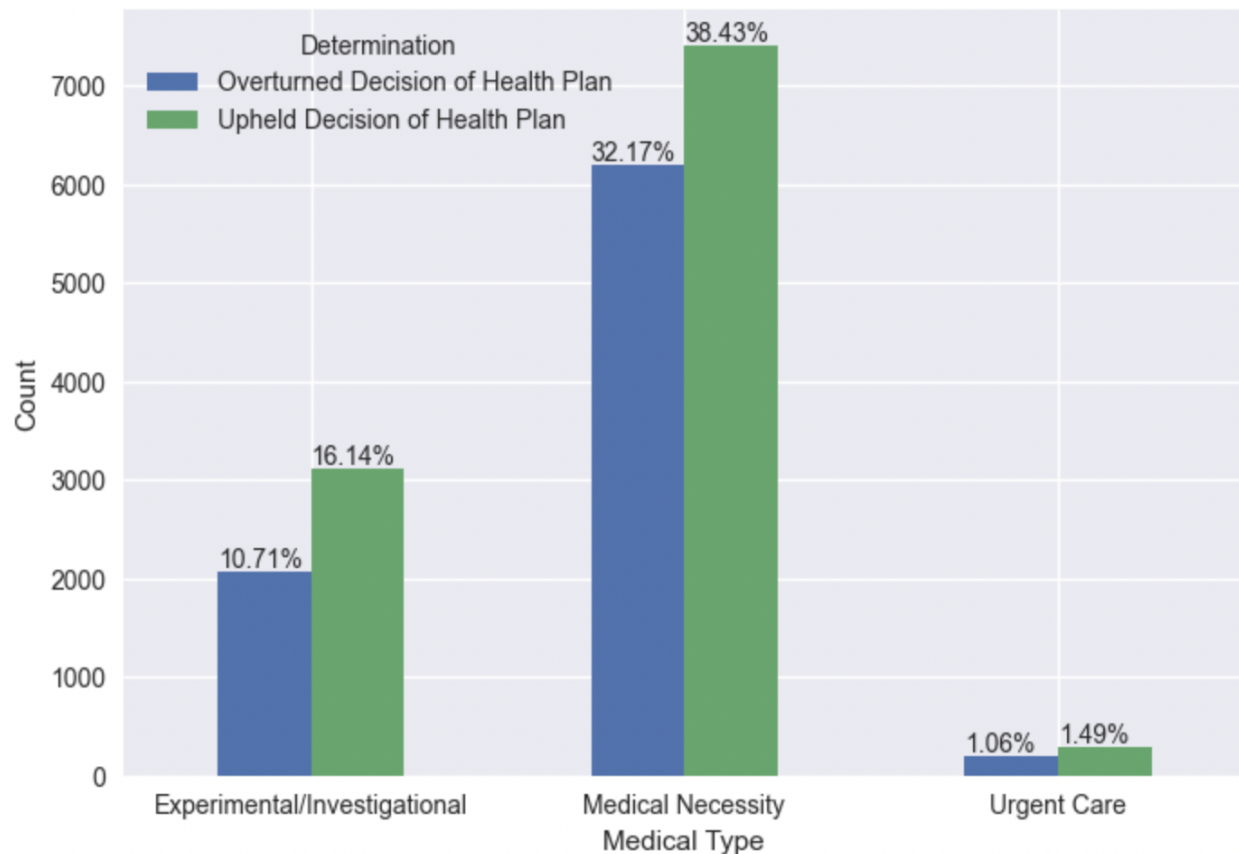
The program is designed to analyze Independent Medical Review (IMR) data provided by the California Department of Managed Health Care (DMHC). The program is designed to identify patterns and trends in the IMR data, which can help to provide insight on what type of cases are overturned or upheld.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Output and Analysis**

Graphical representation of IMR determination outcomes for different types of medical services.



#### **Medical Types:**

- **Medical Necessity:** defined by health insurance plans as treatments, tests, or procedures necessary to maintain or restore health.
- **Experimental/Investigational Type:** services whose safety or efficacy is not known or are services that are used in a way that departs from generally accepted standards of practice in the medical community
- **Urgent Care:** a medical condition that is not potentially life-threatening and doesn't require immediate medical attention but still requires care within 24 hours.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

## **Comparison Questions #2 - Gender Determinations Deep Dive**

Are there any differences in the IMR determination outcomes between Genders for each Diagnosis Category?

- Unit of analysis: Patient's Gender (Male/Female), Determination, Diagnosis Category, and sub-categories.
- Comparison values:
  - Raw counts of cases overturned compared to upheld based on diagnosis category.
  - Percentage of cases overturned compared to upheld based on diagnosis type.
  - Odds of cases being overturned compared between Female and Male.
- Computations:
  - Tree map
  - Bar chart
  - Chi-square test of independence to determine if there's an association between IMR determination and Patient's Gender.
  - Natural language processing

### **Statistical Testing:**

A chi-square test of independence was performed to examine the relation between case Determination and Gender. The relation between these two variables were considered statistically insignificant,  $[X^2(Df=1, N=19236) = 0.35, p > .05]$ . Since our p-value is greater than our significance level set to 5%, we fail to reject the null hypothesis of independence. Therefore, we do not have sufficient evidence to conclude there is a relationship between Gender and the Determination.

### Chi-squared test for independence

Test	Df	Chi-square	P-value
Pearson	4	0.383015	0.983844
Log-likelihood	4	0.383061	0.98384

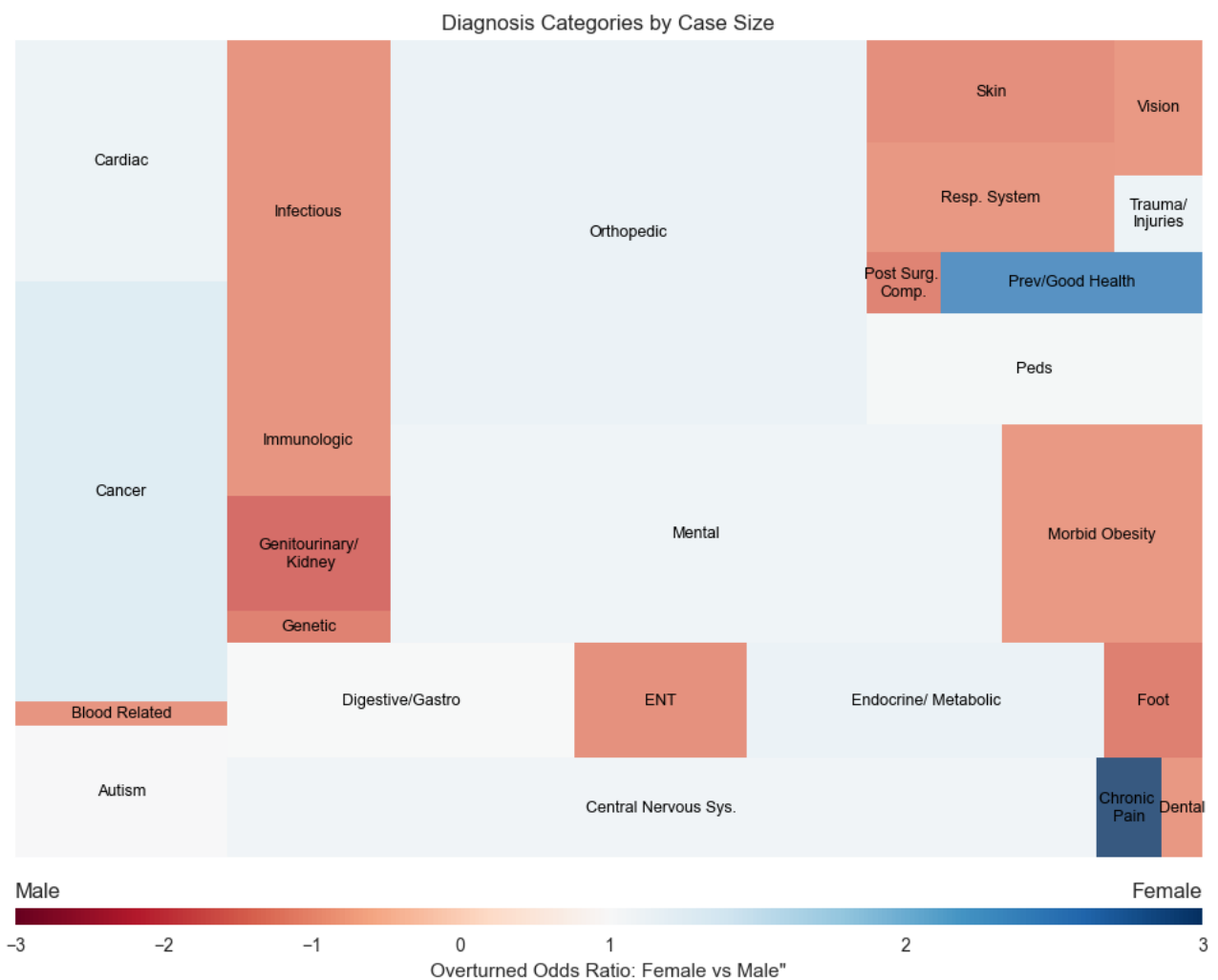
	Overtured	Upheld	Total	Overtured %
<b>Patient Gender</b>				
<b>Female</b>	4914	6218	11132	44.14
<b>Male</b>	3541	4563	8104	43.69

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

by Diagnosis Category

Based on the tree map below, a few specific diagnosis categories tend to favor either female or male. Darker colors represent higher odds, which translates to higher probability of being overturned. The top three categories with the highest bias to a specific gender are Chronic Pain (female), Prevention/Good Health (female), and Genitourinary/Kidney (male). After further analysis



Regarding Prevention/Good Health, a significant proportion (75%) of the reviewed cases focused on mammography, an imaging technique for breast examination. Notably, these cases were specific to females, accounting for 90% of the total. Among the female cases, approximately 66.67% of the mammography-related determinations were overturned. These

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

findings underscore the importance of considering specific factors and outcomes within this category.

In the Genitourinary/Kidney category, the majority of cases were associated with sub-categories like erectile dysfunction and urology, primarily affecting males. Notably, the sub-category of erectile dysfunction witnessed a significantly higher number of overturned cases compared to others within Genitourinary/Kidney. This pattern contributes to the overall favorability towards males in this category, primarily driven by the prevalence of cases related to erectile dysfunction.

Concerning Chronic Pain, an analysis of treatment categories including pharmacy/prescription drugs, rehabilitation services, and pain management remedies revealed no clear patterns or trends in determinations. The cases were scrutinized with a focus on precise patient care, rendering any comparisons statistically insignificant.

For Post-Surgical Complications, it is worth noting that only 18 cases involving males were reviewed. Therefore, any observed differences between genders can be attributed to sampling error, which may arise from factors such as selection bias and potential measurement error due to subjectivity in the documentation process.

In the cases of Orthopedic, Mental Health, and Cancer, when considering treatment categories and subcategories, no discernible biases were identified between genders in determination outcomes. Decisions were primarily driven by the unique circumstances of each patient during the review process.

#### Natural Language Processing:

In order to gain insight into the common themes and issues present in the decisions of IMR, a natural language processing technique was used to extract the most frequent combinations of words between each gender. The top 10 word combinations obtained from this analysis are provided below and are expected to provide valuable information for understanding the types of cases reviewed by IMR. It is important to note that some of the bigrams may overlap, as is the case with "hepatitis," which appears both at number 1 and 10. This suggests that they were likely part of the same sentence, such as "chronic hepatitis c." While this type of analysis is not perfect, it can still provide valuable insight.

The analysis of the most common bigrams for men and women in IMR decisions revealed interesting patterns. For men, the most frequent combinations of words were related to conditions such as hepatitis C, back pain, disc replacement, and prostate cancer. On the other hand, women's cases frequently involved themes of weight loss, breast cancer, back pain, hepatitis C, gastric bypass, and breast tomosynthesis.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

These findings highlight the specific health concerns that are often addressed in IMR decisions for each gender. Hepatitis C and back pain appeared as common issues for both men and women. The presence of certain conditions, such as prostate cancer for men and breast cancer for women, reflects the gender-specific health challenges considered in the review process.

Top 10 most common bigrams in Male patient's findings:

1: hepatitis c Occurrences: 845	6: association study Occurrences: 352
2: growth hormone Occurrences: 478	7: study liver Occurrences: 351
3: back pain Occurrences: 459	8: liver diseases Occurrences: 346
4: patients chronic Occurrences: 407	9: prostate cancer Occurrences: 341
5: disc replacement Occurrences: 378	10: chronic hepatitis Occurrences: 340

Top 10 most common bigrams in Female patient's findings:

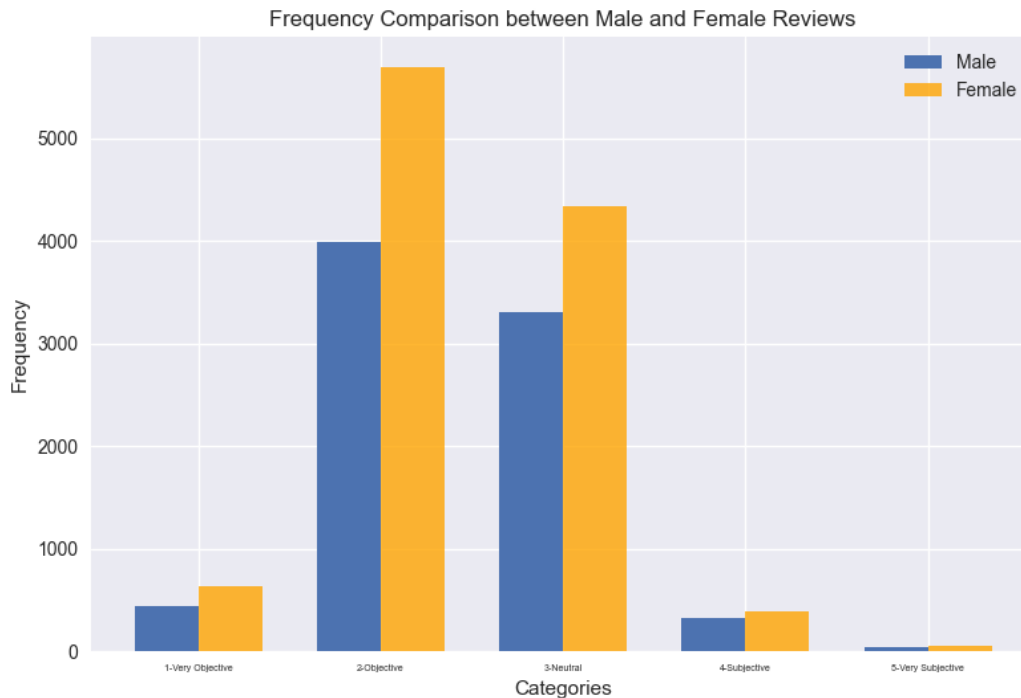
1: weight loss Occurrences: 1189	6: breast tomosynthesis Occurrences: 638
2: breast cancer Occurrences: 1121	7: loss surgery Occurrences: 435
3: back pain Occurrences: 744	8: botox injections Occurrences: 423
4: gastric bypass Occurrences: 682	9: duodenal switch Occurrences: 407
5: hepatitis c Occurrences: 648	10: chronic pain Occurrences: 398

The bar chart below illustrates that the frequencies are consistent across genders. It is important to note that approximately 27.5% of the cases reviewed pertain to women, which explains the higher frequencies observed for females. However, the overall trend remains the same. Irrespective of gender, the vast majority of cases fall under the objective or neutral category. Cases classified as subjective to very subjective only represent a small percentage (4-4.5%) across both genders.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---



Male Reviews %		Female Reviews %	
Level of Subjectivity		Level of Subjectivity	
1-Very Objective	5.392399	1-Very Objective	5.758175
2-Objective	49.284304	2-Objective	51.221703
3-Neutral	40.807009	3-Neutral	38.959756
4-Subjective	4.022705	4-Subjective	3.566295
5-Very Subjective	0.493583	5-Very Subjective	0.494071

## Ryan's Conclusions

Based on our analysis, the relationship between case Determination and Gender was found to be statistically insignificant, indicating no significant association between the two variables. Our findings suggest that determinations were driven more by individual patient cases rather than gender biases. However, specific diagnosis categories showed tendencies favoring either female or male determinations. For instance, Prevention/Good Health cases primarily involved mammography, which was predominantly specific to females and had a relatively high rate of overturning. In Genitourinary/Kidney cases, sub-categories like erectile dysfunction contributed to a higher number of male determinations being overturned. No clear patterns or biases were

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

observed in Chronic Pain, Orthopedic, Mental Health, and Cancer cases when considering treatment categories and subcategories. It is important to note that the limited number of reviewed Post-Surgical Complications cases involving males may introduce sampling error. Overall, our analysis suggests that while some diagnosis categories exhibited gender-related biases, individual patient cases played a significant role in determination outcomes.

In addition to the previous analysis, it is important to note that women accounted for 37.3% more cases than men in our dataset. Even when accounting for obstetrics and gynecology (OBGYN) cases, which are specific to women, there was still a 27.5% higher proportion of women's cases. It is crucial to emphasize that the cases reviewed were health plan denials, and while no biases were noted in the reviewers' determinations, the higher rate of cases where a women's health plan denies coverage indicates a notable trend.

These findings shed light on potential disparities in health plan coverage for women, highlighting the need for further investigation into the reasons behind these discrepancies. Understanding the factors contributing to a higher rate of health plan denials for women can help identify areas for improvement and ensure equitable access to healthcare services.

## **Bill's Analysis**

### **Comparison Question #2**

**Are there any differences/apparent biases in determinations by the IMR board based on gender? Or age group?**

- Unit of analysis: Gender and Determination
- Unit of analysis: Age Grouping and Determination
- Comparison values:
  - Raw counts of cases overturned compared to upheld based on gender
  - Percentage of cases overturned (positive result) for males vs. females
  - Raw counts of cases overturned/upheld based on age group
- Computations/visualizations:
  - Statistics
  - Bar chart visualizations

## **Brief Description of the Program**

This program (available in our Jupyter notebook) analyzes the following data within our dataset:

Team:

Ryan Summers, Bill Steel, Andrew Alford

---

- **Determination:** Either “overturned” or “upheld.” The favorable result for the patient is that the IMR board overturns the decision of the Insurance Company to not provide the service.
- **Gender:** Categorized as Male or Female by the IMR
- **Age Range:** Age grouping of the patients. 0-10, 11-20, 21-30, 31-40, 41-50, 51-64, 65+

The program starts by 1st looking at the total of upheld vs. overturned decisions and plots a bar chart showing the results. The second portion of the program breaks down the data by gender and generates the statistics as well as a bar chart to visualize the results. The third portion groups the data by age range and generates statistics as well as a bar chart for visualization.

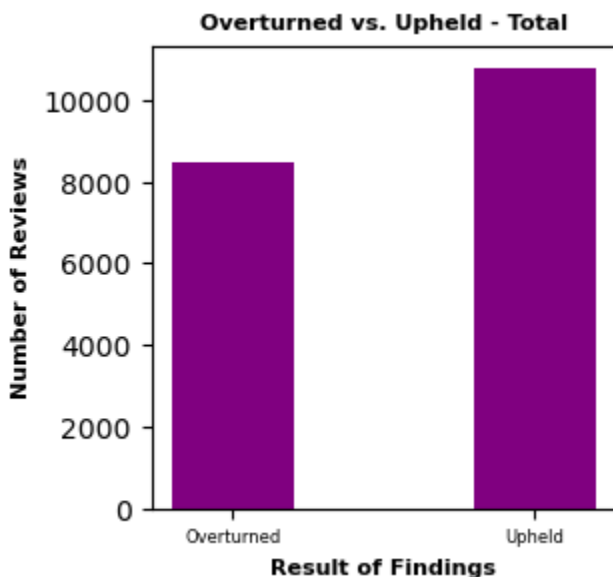
Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

## **Output and Analysis**

### **Overall:**

Out of 19245 reviews, 8457 were overturned while 10788 were upheld which reflects an overall 43.94% success rate for the patient. The obvious general conclusion here is that ~56% of the time the insurance company decisions are upheld. That said, generally speaking, there certainly is a reasonable chance that the decision will be overturned.



### **By Gender:**

What's clear from the analysis is that female patients have significantly more claims than male patients (57.88% vs. 42.12%). What is also clear is that there does not appear to be any bias based on gender in terms of overturning the decision by the insurance company. Decisions related to female patients were overturned 44.14% of the time. While decisions for male patients were overturned 43.70% of the time.

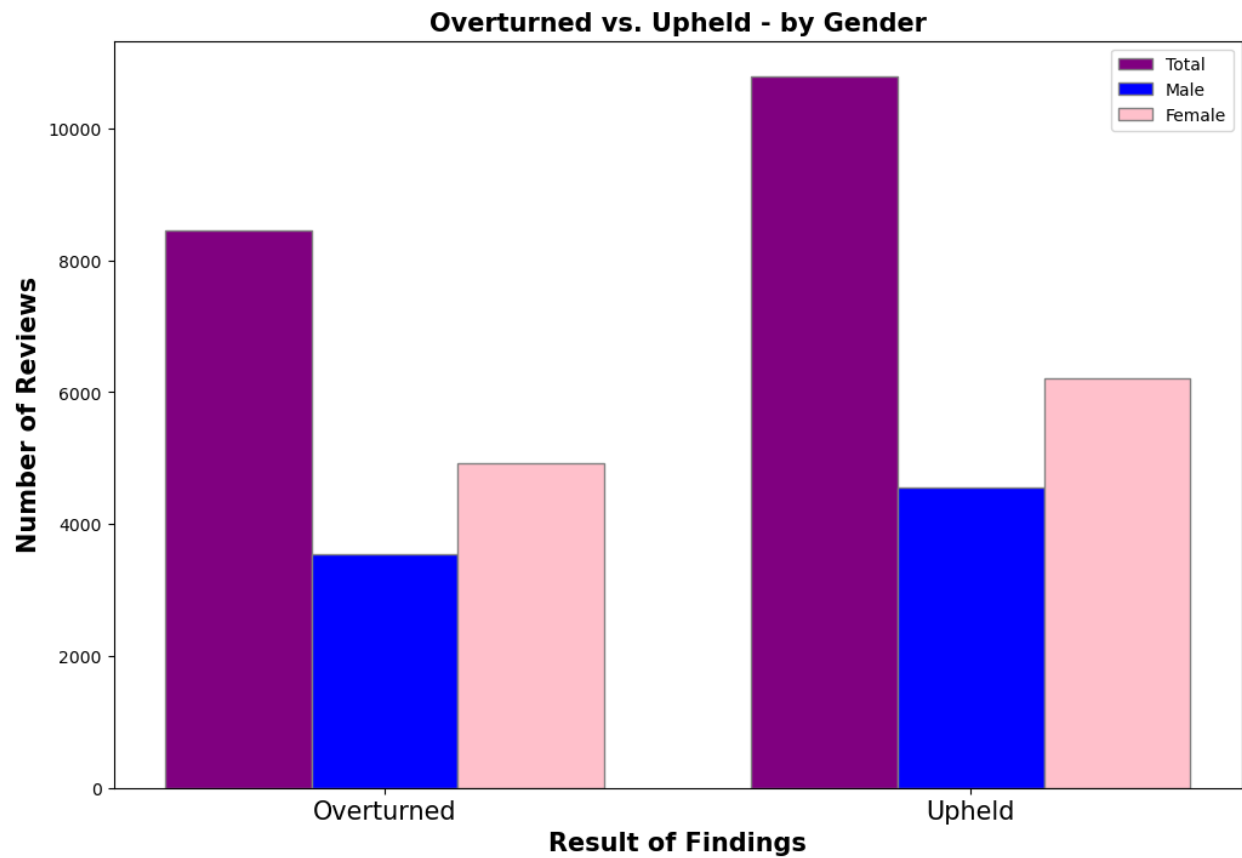
### **Raw Statistics:**

Male: 8103, Female: 11133, missing/other: 9  
Percent of claims related to Males 42.12%  
Percent of claims related to Females 57.88%  
total records: 19245  
Total valid records = 19236  
Male Overturned 3541  
Male Upheld 4562  
Female Overturned 4914  
Female Upheld 6219  
Female Overturned 44.14%  
Male Overturned 43.70%

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

Visual representation of the results:

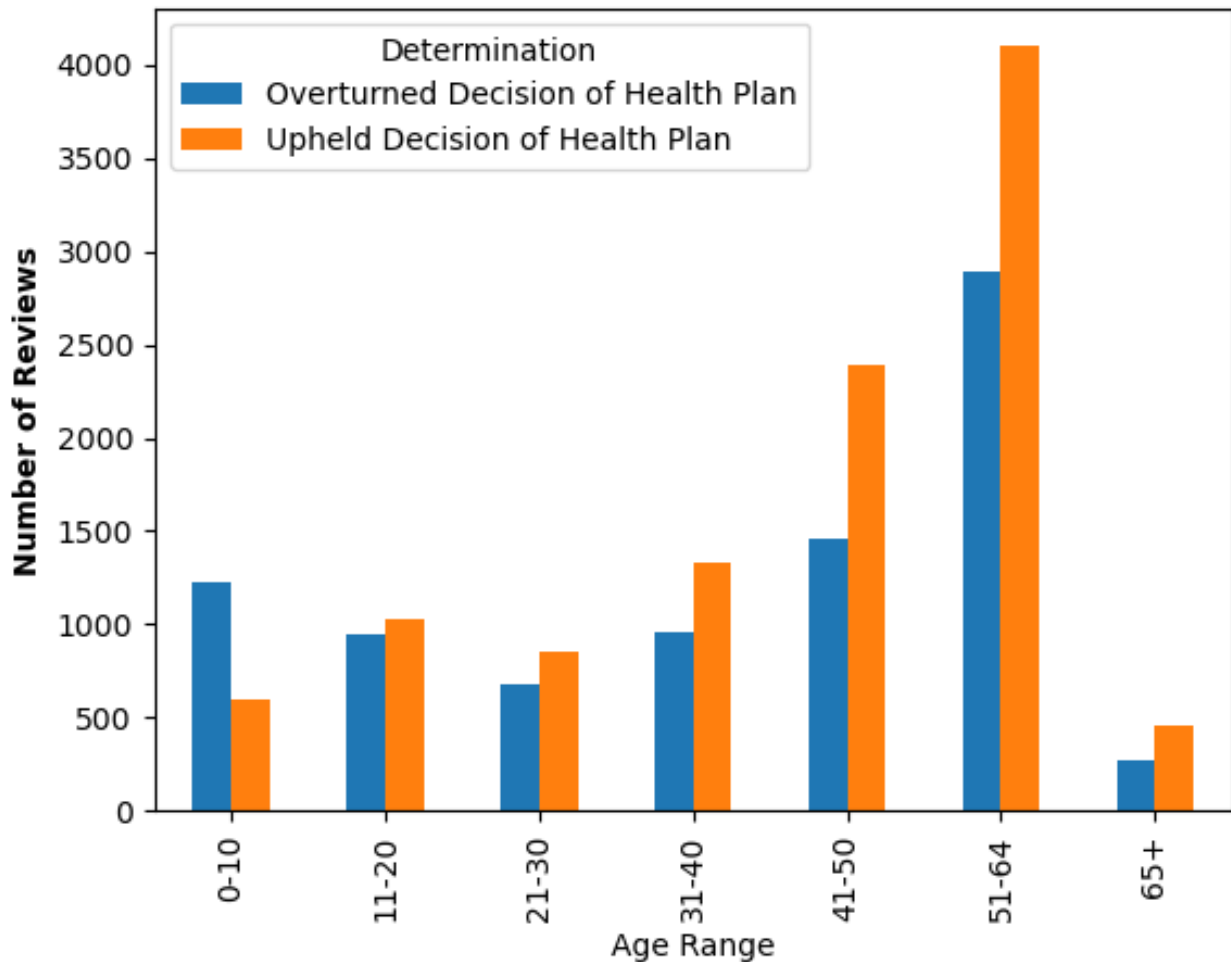


Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

By Age Group:

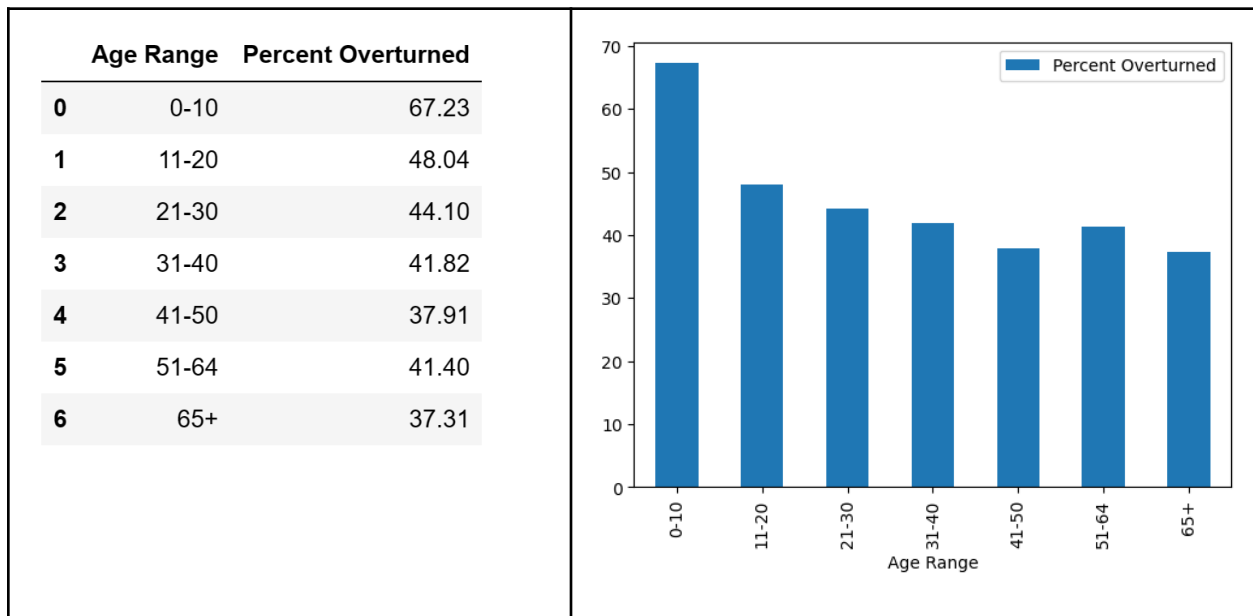
Based on age group, as per the chart below, it's clear that the vast majority of patient reviews occur within the 51 to 64 year old age group followed by the 41 to 50 year old age group. Interestingly, the 65+ age group has comparatively minimal IMR reviews which suggests that the insurance companies typically approve services for patients in this age range.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

Also quite noteworthy based on the previous visual is how often the insurance companies are overturned for children between the ages of 0-10 and how often the insurance companies' decisions are upheld for anyone 40 and older. This is made more obvious based on the data and visual shown below. It's quite astounding to see a variance from 67.23% for 0-10 year old patients vs. a 37.31% for 65+ year old seniors.



### **Bill's Conclusions**

1. Generally, insurance company decisions are upheld (56% of the time). That said, there certainly is a reasonable chance that the decision will be overturned (44%).
1. Female patients have significantly more claims than male patients (57.88% vs. 42.12%).
2. There is no bias based on gender in terms of overturning the decision by the insurance companies. Decisions related to female patients were overturned at about the same rate as men (~44%).
3. The vast majority of patient reviews occur within the 51 to 64 year old age group followed by the 41 to 50 year old age group.
4. The 65+ age group has a remarkably minimal number of IMR reviews. This suggests that the insurance companies typically approve services for patients in this age range and therefore little IMR reviews are needed.
5. Any denial of service for children between the ages of 0-10 should be challenged as these decisions have been overturned over 67% of the time.
6. Remarkably, if you are senior (65+), you have the lowest rate of a successful challenge (37%). As mentioned in point 5, this suggests that the insurance companies typically approve services in the first place for these patients making a challenge less likely and less likely to be overturned.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Comparison Questions #3**

**Is there a relationship between the outcome of an IMR determination and the medical specialty of the physician who made the determination?**

- Unit of analysis: Diagnosis Category and Determination
- Comparison values:
  - Raw counts of cases overturned/upheld based Physician Type (Diagnosis Category)
- Computations/visualizations:
  - Statistics displayed within a bar chart visualization

### **Brief Description of the Program**

This program (available in our Jupyter notebook) analyzes the following data within our dataset:

- **Determination:** Either “overturned” or “upheld.” The favorable result for the patient is that the IMR board overturns the decision of the Insurance Company to not provide the service.
- **Diagnosis Category:** Diagnosis Category refers to the type of Physician that is the subject matter expert concerning the IMR review.

The program leverages grouping within the IMR dataset (19245 records) in order to understand the number of determinations (overturned vs. upheld) related to the Diagnosis Category.

### **Output and Analysis**

Out of 29 diagnosis categories representing the type of physician, there are a few interesting findings. The largest of which is the number of IMR reviews orthopedic/musculoskeletal (~3500). Additionally, ~2500 of these reviews ruled in favor of the insurance companies. The sheer number of reviews in this diagnosis category and the overwhelming upholding of the insurance plans is visually obvious as per the subsequent bar chart.

Also noteworthy are 3 diagnosis categories where the overturned rate is more than 2 times that of the rate of upholding the insurance companies’ decisions; these being Autism, Infectious diseases, and Pediatrics.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

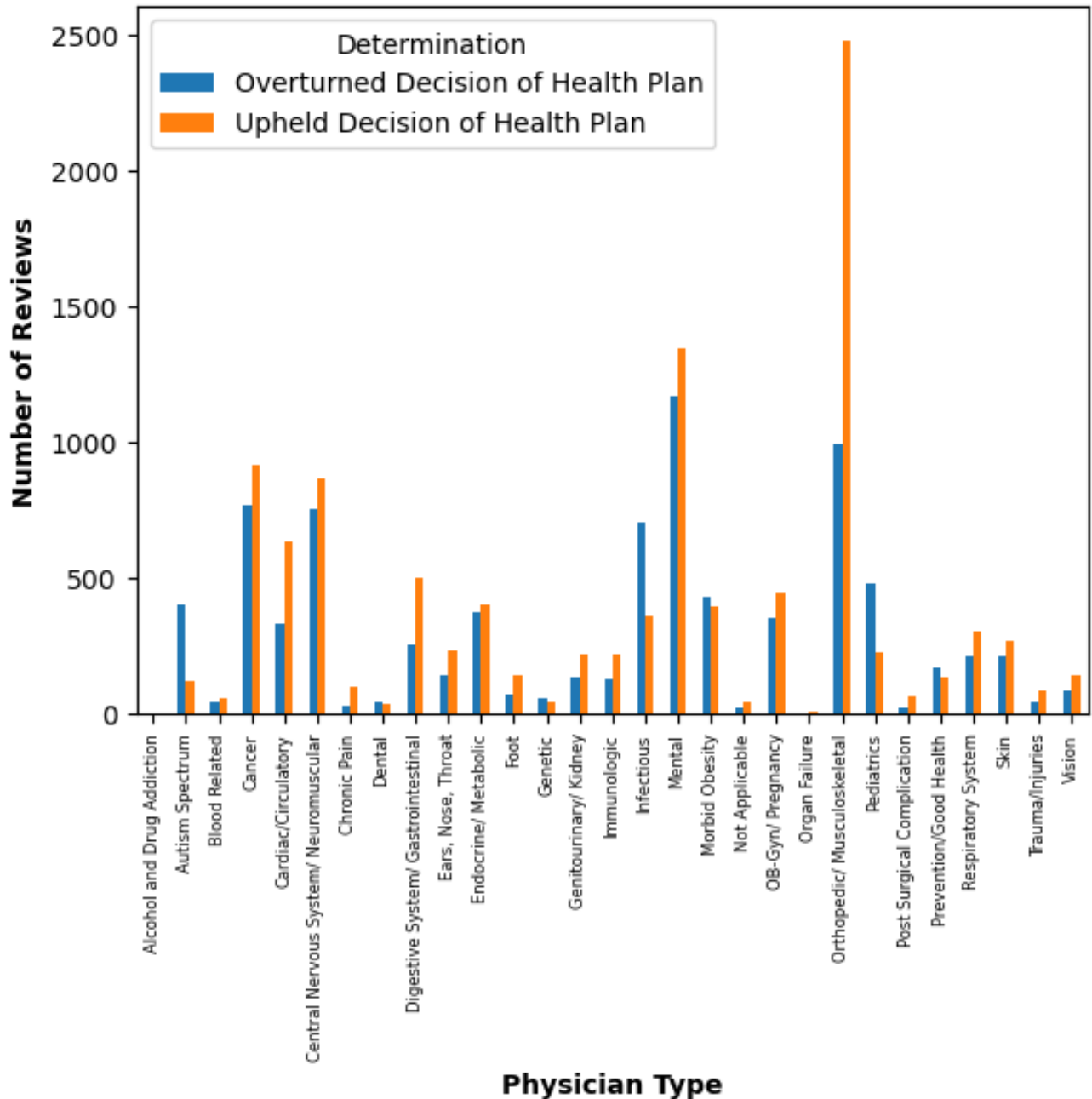
Physician Type	Percent Overturned	Total Reviews	Notes:
Alcohol and Drug Addiction	33.33	3	<i>Alcohol and Drug Addiction has virtually no reviews.</i>
Autism Spectrum	77.10	524	
Blood Related	43.56	101	
Cancer	45.75	1681	<i>Autism has the highest rate of overturning the insurance companies.</i>
Cardiac/Circulatory	34.61	965	
Central Nervous System/ Neuromuscular	46.48	1620	
Chronic Pain	22.13	122	<i>Both "Infectious" and "Pediatrics" also have high overturn rates (66.29% and 67.84% respectively).</i>
Dental	57.89	76	
Digestive System/ Gastrointestinal	33.64	758	
Ears, Nose, Throat	38.03	376	<i>While "chronic pain" and "Post Surgical Complication" have the lowest overturn rates (22.13 and 22.35% respectively), the number of reviews is relatively low in both cases. The significant finding is around "Orthopedic/ Musculoskeletal" with the most reviews and the lowest overturn rate (28.54%).</i>
Endocrine/ Metabolic	48.14	779	
Foot	33.64	214	
Genetic	55.56	99	
Genitourinary/ Kidney	38.29	350	
Immunologic	37.03	343	
Infectious	66.29	1059	
Mental	46.42	2512	
Morbid Obesity	52.31	824	
Not Applicable	38.10	63	
OB-Gyn/ Pregnancy	44.19	801	
Organ Failure	37.50	8	
Orthopedic/ Musculoskeletal	28.54	3469	
Pediatrics	67.84	709	
Post Surgical Complication	22.35	85	
Prevention/Good Health	56.33	300	
Respiratory System	41.33	513	
Skin	44.07	481	
Trauma/Injuries	34.65	127	
Vision	38.39	224	

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### Determinations vs Physician Type

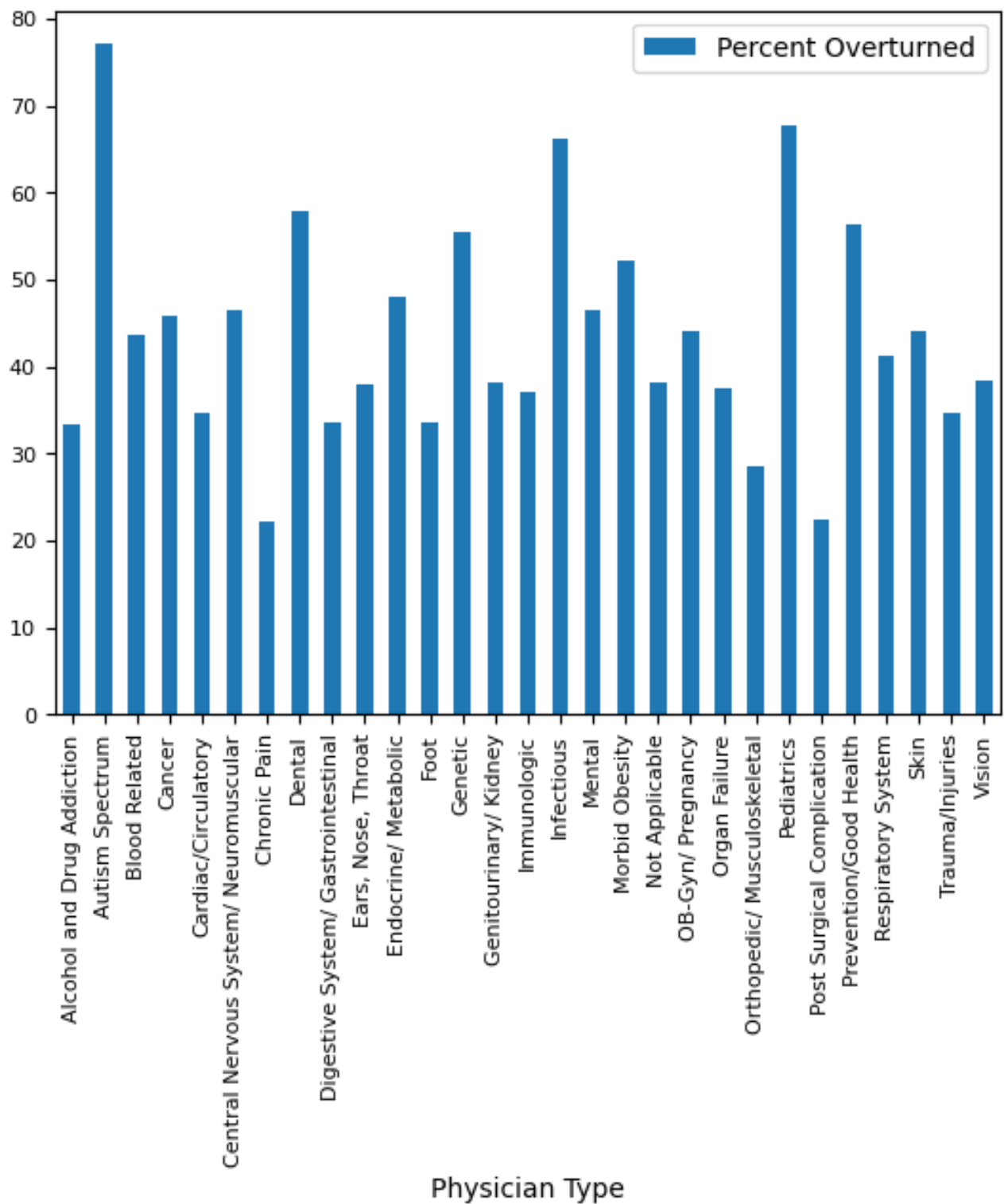
The bar chart below visually shows the key points previously discussed.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

Percent Overturned by Physician Type



Team:  
Ryan Summers, Bill Steel, Andrew Alford

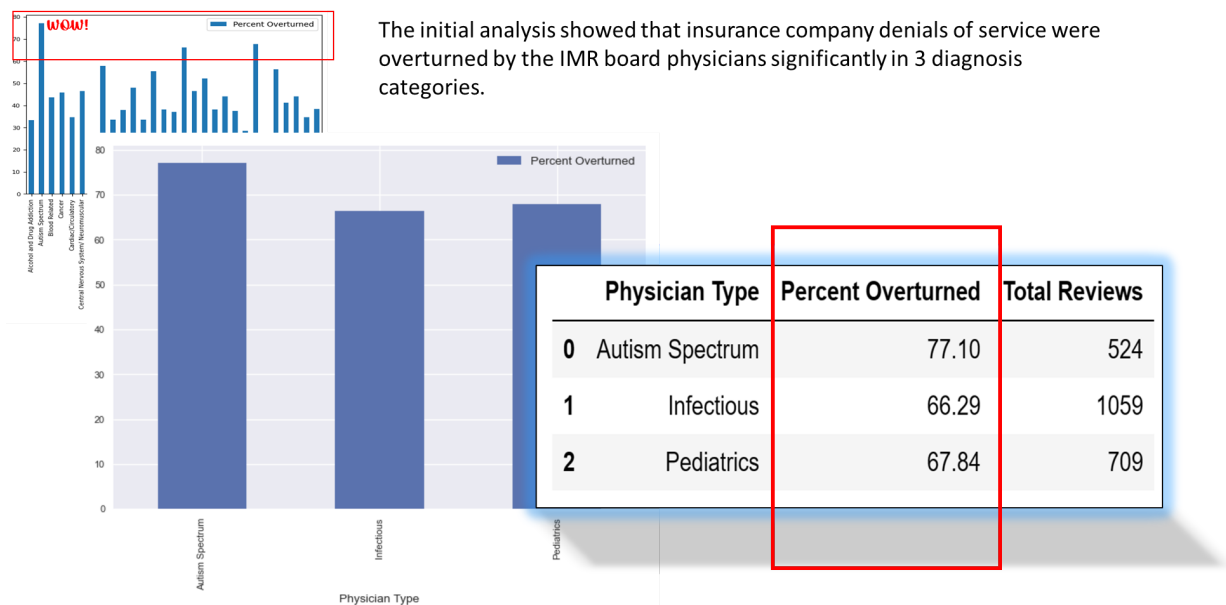
---

**Bill's further analysis on these conclusion (deep dive):**

**Original conclusion:** 3 diagnosis categories are commonly overturned with a rate that is more than 2 times that of the rate of upholding the insurance companies' decisions; these being Autism, Infectious diseases, and Pediatrics.

**Follow up question:** Is there anything by year that has skewed the Physician type results of the 3 diagnosis categories that have a substantially higher overturn rate than all others? Or has this been generally the same year-by-year?

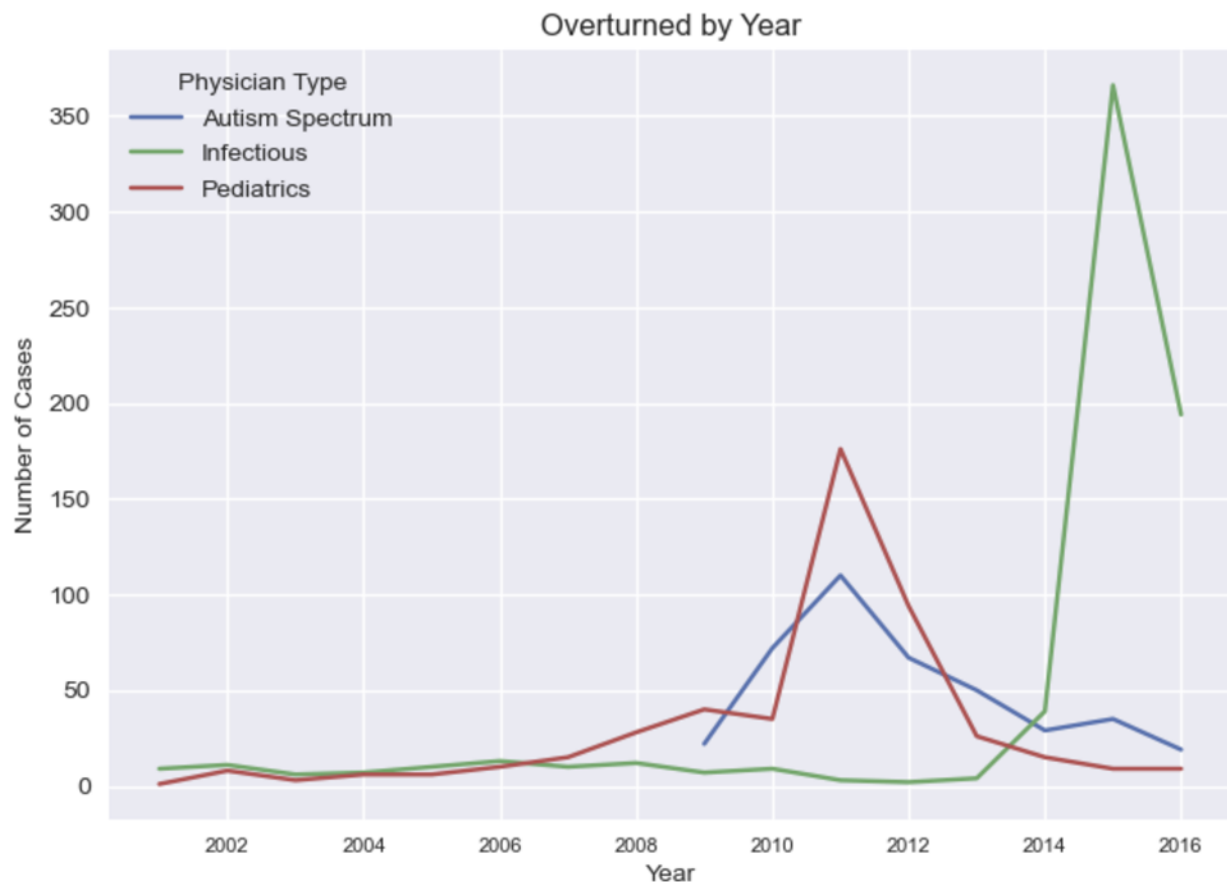
**Subsequent findings:**



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

Looking at the overturned results by year for these 3 categories clearly shows a dramatic increase in certain years. Autism and Pediatric IMR overturned reviews were significantly higher in 2011 thus contributing significantly to the overall results. Infectious IMR overturned reviews were astronomically higher in 2015/2016 driving the overall results. Subsequent analysis could be to understand if any laws were changed in these time periods requiring insurance companies to be more amenable to enrollee claims for these conditions.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

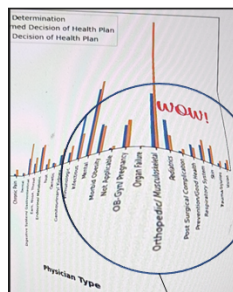
---

**Bill's further analysis on these conclusion (deep dive continued):**

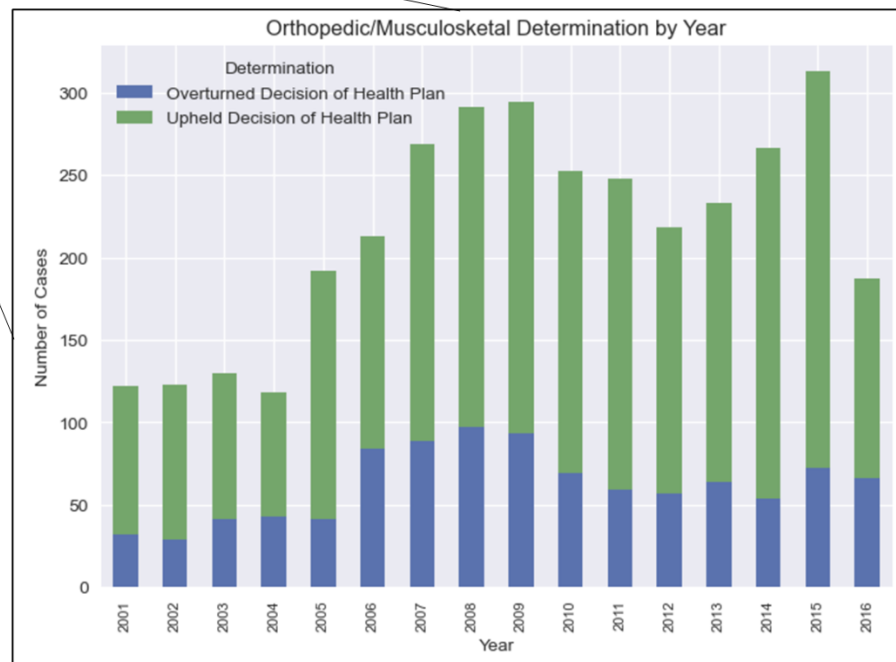
Original conclusion: The largest number of IMR reviews is by far within the orthopedic/musculoskeletal category (~3500). It also has one of the lowest overturn rates (only 28.54%).

Follow up question: Is there anything by year that has skewed the results regarding orthopedic/musculoskeletal remarkably high volume of IMR reviews and low overturn rate? Or has this been generally the same year-by-year?

**Subsequent findings:**



The deeper dive into the Orthopedic/Musculoskeletal analysis by year shows a step change in reviews after 2004 with a significant number of reviews in 2007-2009 and 2014-2015. Generally, the percent overturned is consistent by year without any specific year(s) contributing to the overall low rate of overturned reviews.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Bill's Conclusions**

1. The largest number of IMR reviews is by far within the orthopedic/musculoskeletal category (~3500). It also has one of the lowest overturn rates (only 28.54%).
  - a. The deeper dive into the orthopedic/musculoskeletal analysis by year shows a step change in reviews after 2004 with a significant number of reviews in 2007-2009 and 2014-2015. Generally, the percent overturned is consistent by year without any specific year(s) contributing to the overall low rate of overturned reviews.
2. 3 diagnosis categories are commonly overturned with a rate that is more than 2 times that of the rate of upholding the insurance companies' decisions; these being Autism, Infectious diseases, and Pediatrics.
  - a. The deeper dive into Autism and Pediatric IMR overturned reviews were significantly higher in 2011 thus contributing significantly to the overall results. Infectious IMR overturned reviews were astronomically higher in 2015/2016 driving the overall results. Subsequent analysis could be to understand if any laws were changed in these time periods requiring insurance companies to be more amenable to enrollee claims for these conditions.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

## **Andrew's Analysis**

### **Comparison Questions #5**

**What is the distribution of outcomes for the IMR determinations? What does the summarization of data look like when visualized compared to original expectations?**

- Summarization of data as presented in clean dataset.
- Unit of analysis: Year and Determination.
- Comparison values:
  - Raw counts of cases overturned compared to upheld based on year.
- Computations/visualizations:
  - Vertical bar, horizontal bar, and pie charts.

### **Brief Description of the Program**

This program (available in our Jupyter notebook) analyzes the following data within our dataset:

- **Determination:** Either “overturned” or “upheld.” The favorable result for the patient is that the IMR board overturns the decision of the Insurance Company to not provide the service.
- **Year:** The study encompasses sixteen years from 2001 to 2016.
- **Cleaned Dataset Columns:** Each column has a visualization that displays its data in an easily digestible manner.

The program starts by taking the determinations and plots them onto a bar chart on a by year basis. Instead of simply showing a total, it breaks it down by either overturned or upheld decisions. The second portion of the program shows a visualization of each column to break down how each factor within the dataset is portrayed.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

## **Output and Analysis**

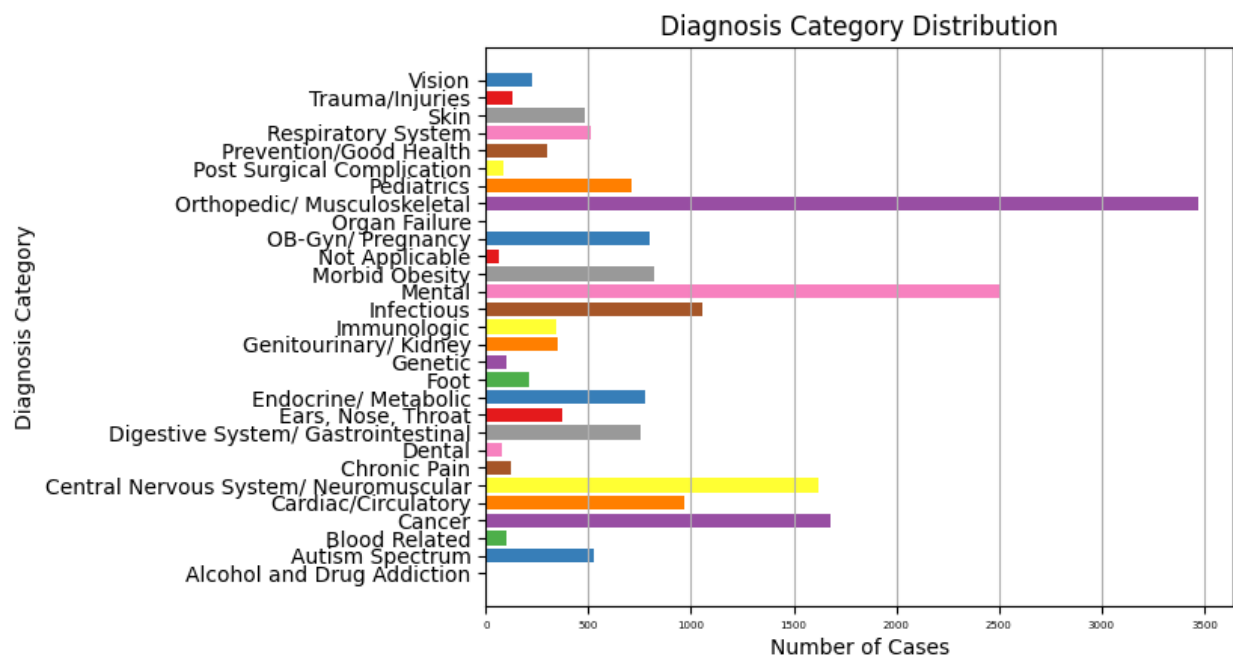
### **Dataset Summarization:**

Using string formatting, I created a paragraph that summarized the basic parameters of the dataset's columns. The paragraph is as follows: "The dataset has 15 columns and 19245 rows. It uses metrics from 16 years, 30 diagnosis categories, 279 diagnosis sub categories, 32 treatment categories, 284 treatment sub categories, whether or not the determination was upheld or overturned, 3 determination types, 8 age ranges, 3 patient genders, and a report on findings for each reference ID."

Though there are more charts that summarize the columns, in the interest of space and time I have included only the most unique and interesting plots in this report. For a look at all the charts, please refer to the Jupyter notebook attached to the submission.

### **Diagnosis Category:**

This plot shows that there are four distinct outliers when it comes to the distribution of the diagnosis categories. These are Orthopedic/Musculoskeletal, Mental, Central Nervous System/Neuromuscular, and Cancer. It's also interesting to note that Alcohol and Drug Addiction diagnoses as well as Organ Failure are negligible.

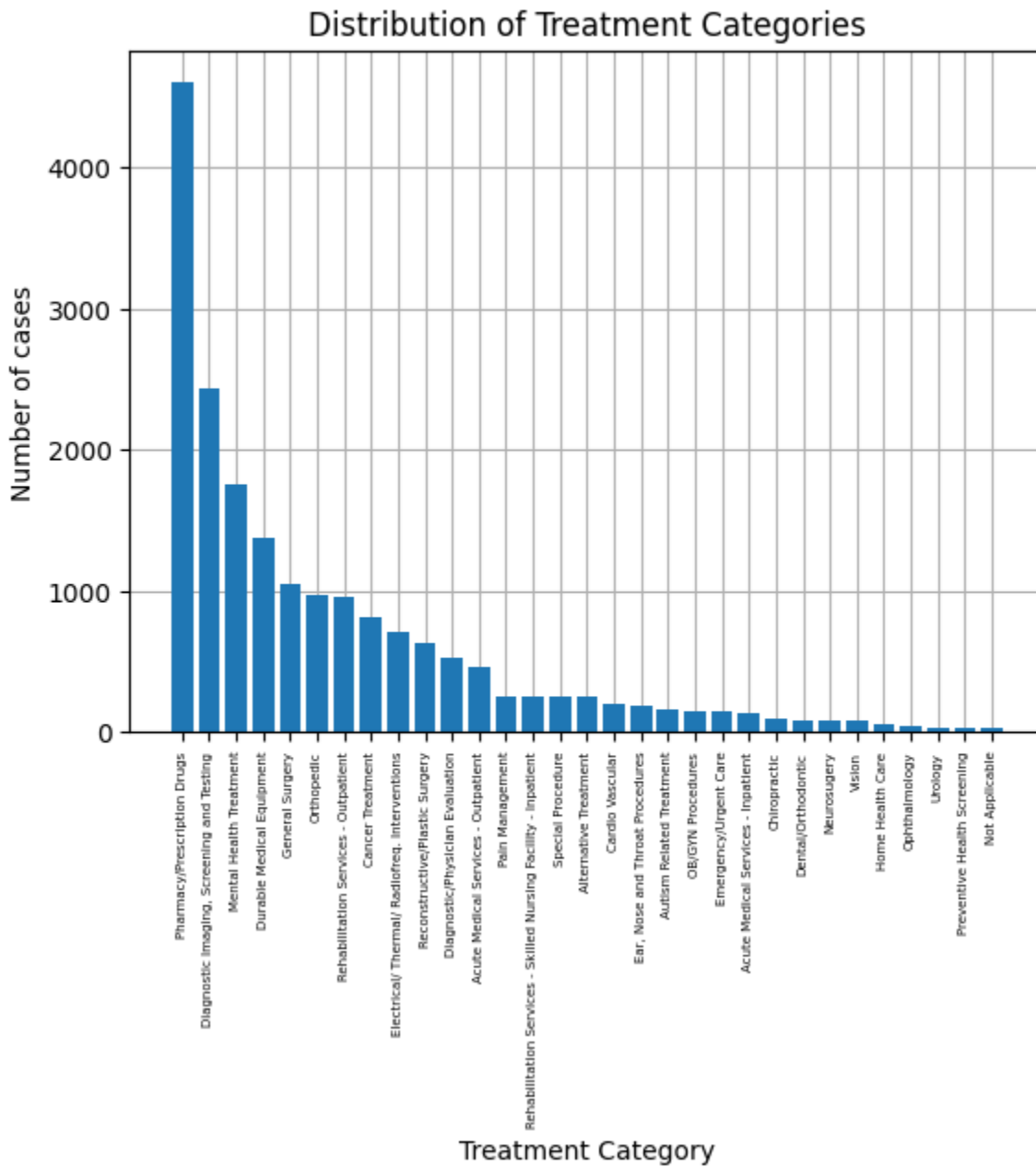


### **Treatment Category:**

This chart shows that there is one key treatment category that far outweighs many of the others, which is clearly Pharmacy/Prescription Drugs. After which there is a steep decline to second place of Diagnostic Imaging. From there it is a slow and steady decline for each category until the category is only a rare diagnostic strategy.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

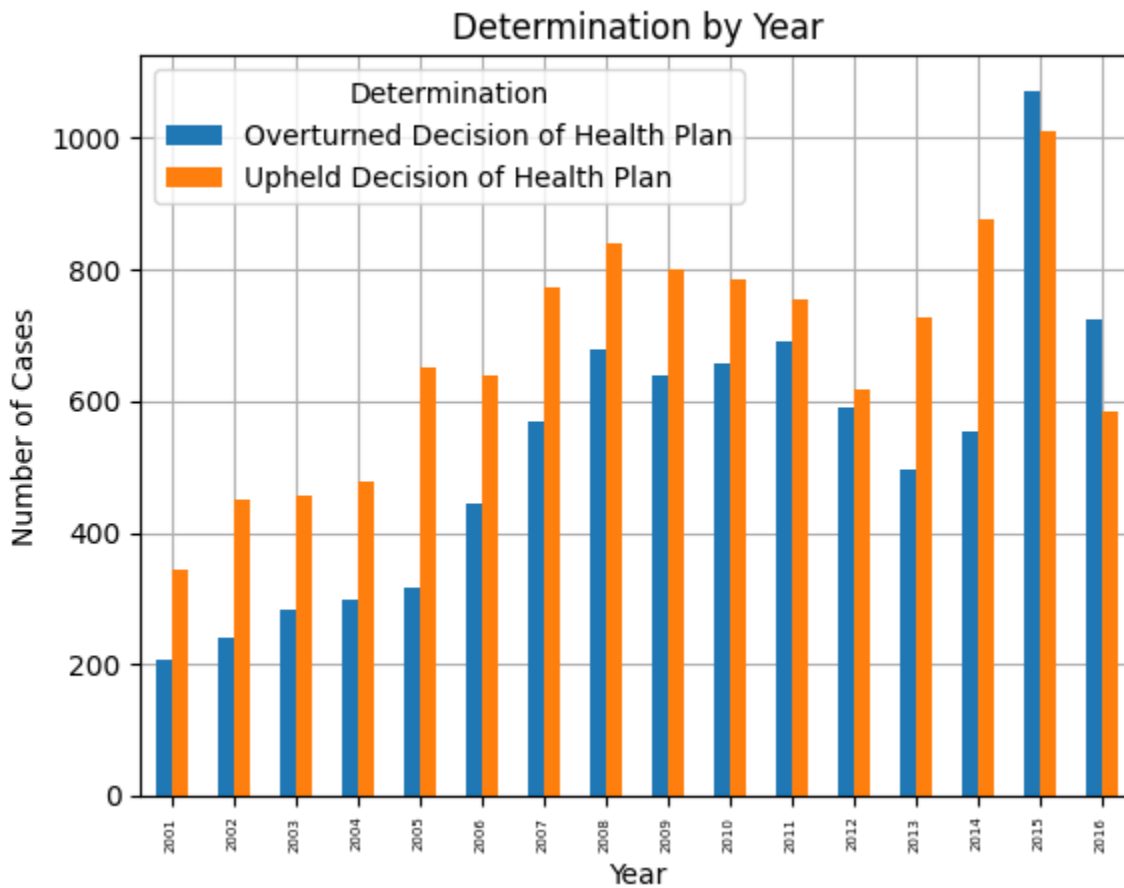


Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

By Year:

From 2001 to 2014, IMRs consistently upheld the decisions of health plans. In 2015, not only was there a significant spike in overall cases, but for the first time in the dataset, overturned decisions outweighed upheld decisions, though not by a massive margin. 2016 only showed a more pronounced continuation of the trend of 2015, with a much higher margin of decisions being overturned compared to upheld.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Andrew's Conclusions:**

1. There are clearly diagnostic categories that are far more frequent than other types, with the most popular being Orthopedic/Musculoskeletal, Mental, Central Nervous System/Neuromuscular, and Cancer.
2. Pharmaceutical treatments are far and away the most frequent recommendations, with nearly 25% of the entire dataset having it being their treatment recommendation.
3. Determinations to overturn or uphold the decision of the health plan remained largely consistent from the years of 2001 to 2014, with a steady increase throughout the spread. This changed in 2015 when there was a sharp spike in cases as well as overturned decisions outweighing upheld decisions for the first time. This trend was more pronounced in 2016, albeit with less cases than 2015.

### **Andrew's further analysis on these conclusions (deep dive)**

*Original conclusion:* Out of all of the different diagnosis categories, there were four major outliers that stood out from the rest. These five included Orthopedic/Musculoskeletal, Mental, Cancer, and Central Nervous System/Neuromuscular. Each of us decided to look more in depth at the top three of these diagnosis categories and see what insights we could find.

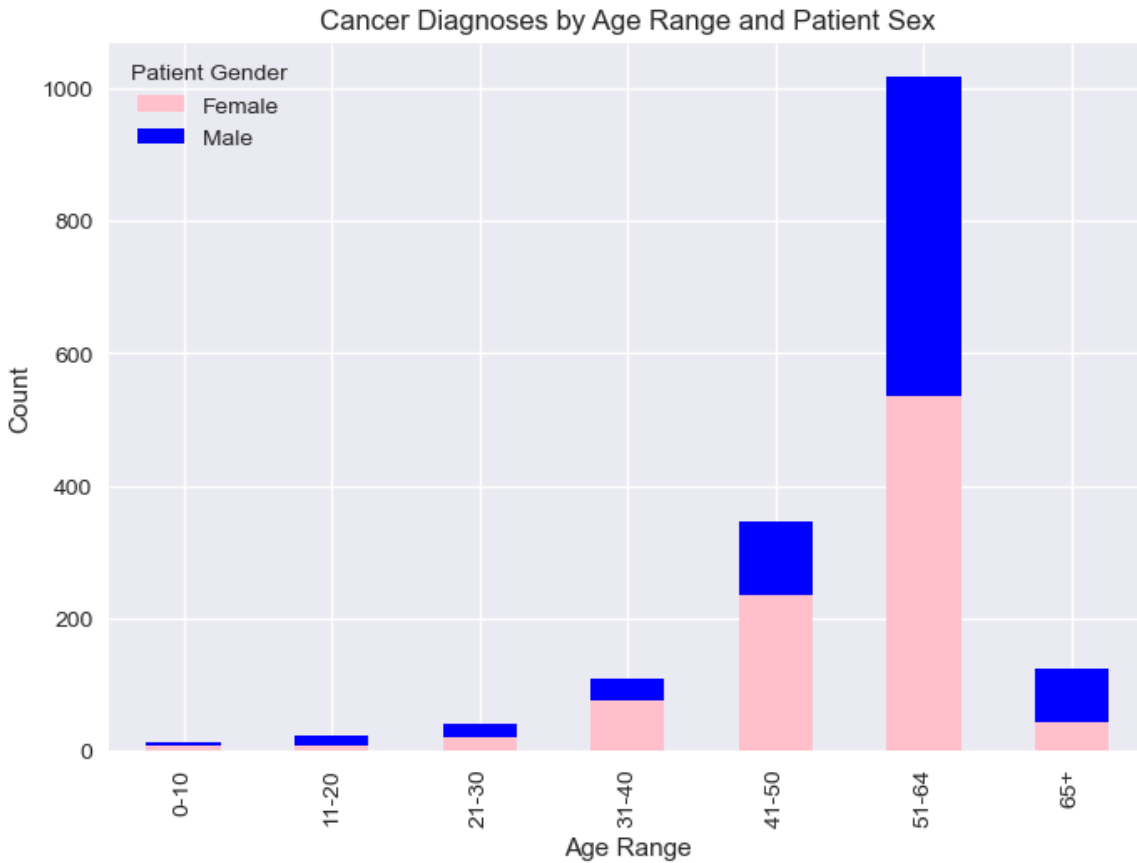
*Follow up question:* Why is it that cancer has such a large number of cases? What are the different types of cancer that come up the most? What are some of the most common treatments that exist for cancer?

### **Subsequent findings:**

As shown in this visualization, cancer diagnoses go up significantly as the age of the patient increases. It starts to increase at the age group of 31-40, significantly increases from 41-50, and then almost triples at the age group of 51-64. Perhaps morbidly enough, there is a sharp decline in cases in the 65+ age group, indicating that many of the patrons who experience cancer do not make it to 65+

Team:  
Ryan Summers, Bill Steel, Andrew Alford

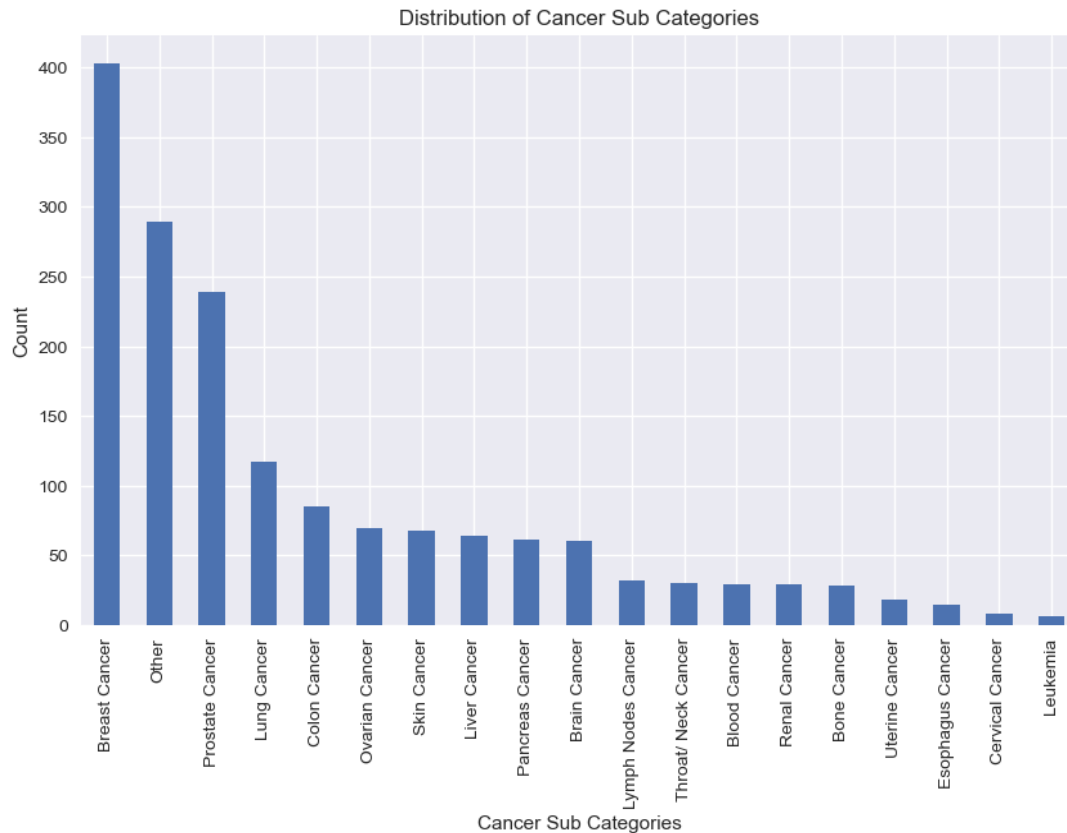
---



Diving deeper into the cancer analysis, I broke down the frequency of cancers by sub category to see which ones were the most prevalent in the data set. It was clear that there were three strong outliers, after which the rates leveled off at a pretty steady rate. The top in the list was breast cancer, with a little over four hundred reported cases. The second was, unhelpfully, 'other' which does not give us much information unfortunately. The third was prostate cancer at a little under two hundred and fifty cases.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

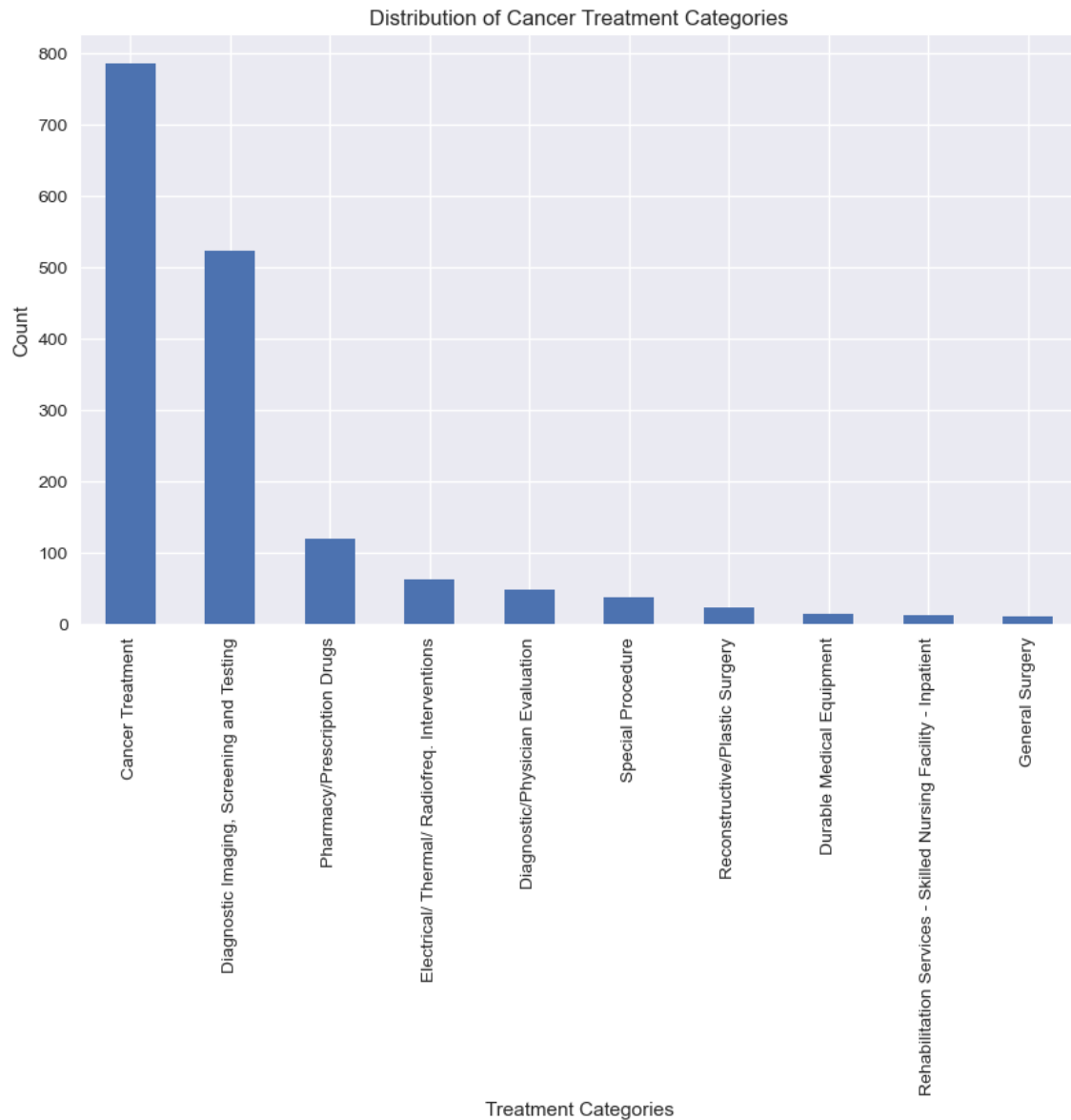
---



At this point, I wanted to answer the third question in the follow up question list, which was what the most common cancer treatment categories are. Unhelpfully, the most common cancer treatment was 'cancer treatment', which provided no detail whatsoever. It could only be guesswork as to what exactly 'cancer treatment' consists of, but it does not help understanding anything. The second highest treatment was diagnostic imaging, screening, and testing, which makes sense for cancer. The treatments dropped significantly after that.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

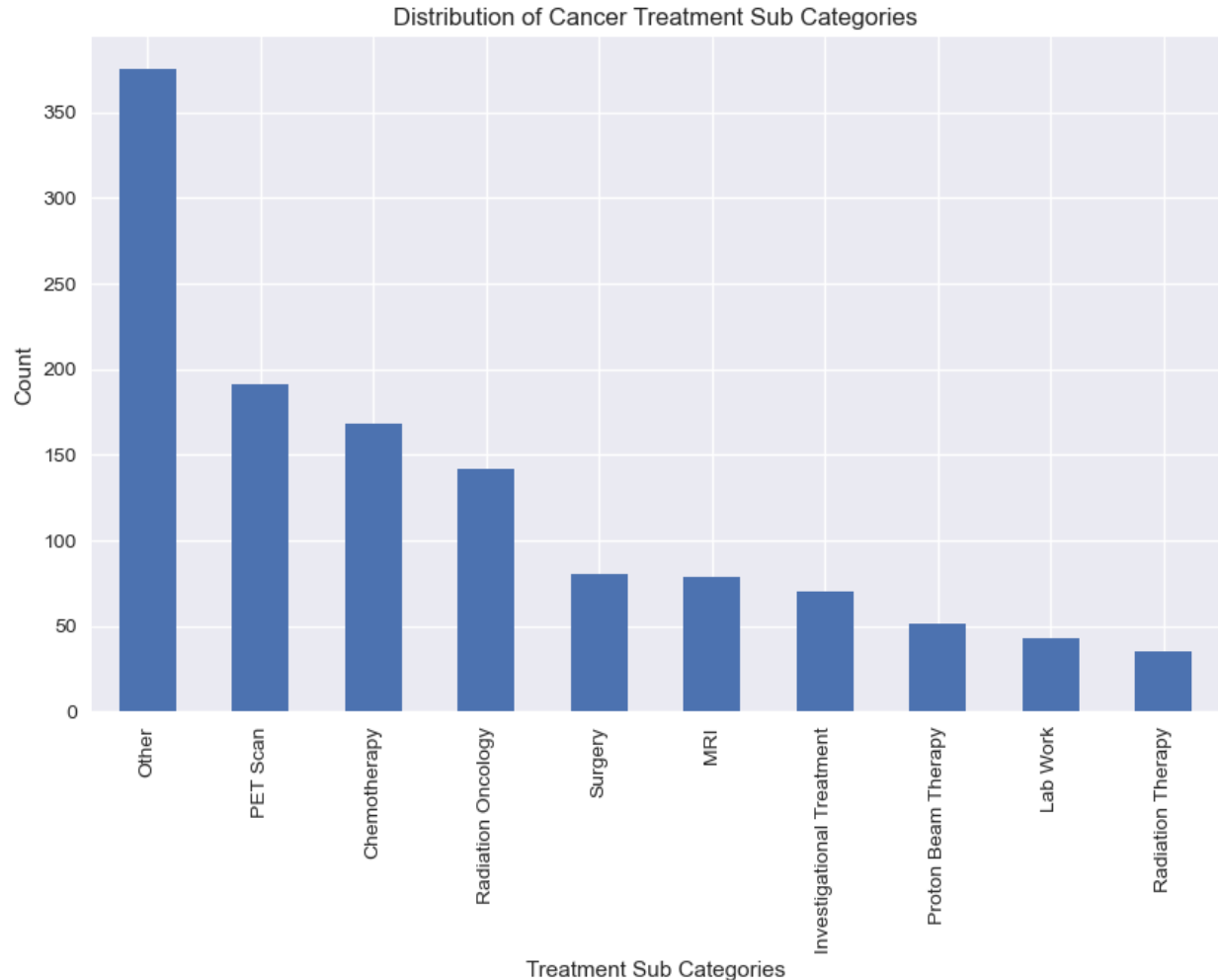
---



Lastly, I decided to go one step further and look at the cancer treatment subcategories, which essentially is the same as the treatment categories, but one step further in detail. Once again, I was left somewhat disappointed, with the most common subcategory treatment being 'other', again giving no detail. However, there were other valuable insights outside of this preliminary result, including PET Scan, Chemotherapy, and other interesting treatments such as Proton Beam Therapy.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---



### **Andrew's Conclusions:**

The amount of conclusions that I was able to gleam from my deep dive analysis are unfortunately still a little unclear given the lack of detail that was put into many of the rows. By this, I mean that the most frequent cancer treatment given was simply 'cancer treatment', and furthermore the most common sub category treatment was 'other'. Despite this, there were many valuable insights that I was able to decipher from the data. The first interesting insight was that recorded cancer rates fell steeply once patients reached the age group 65+, concluding that either a sizable portion died or stopped being recorded due to enrolling in medicare. Secondly, rates of cancer were generally equal between men and women, with women being a slightly greater portion of the rates. The most prevalent form of cancer was breast cancer followed closely by prostate cancer. Lastly, PET scans and chemotherapy were the most common forms of cancer treatment.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Sentiment Analysis:**

A great deal of information was gleaned from the structured data. This section focuses on the unstructured text explanations of the IMR board to understand if there are any biases and/or emotional views based on their 19,245 decisions. These explanations are free form text consisting of ~300 words per review (~6MM words across the entire data set). This is the equivalent of about 18 novels at 1000 pages each.

### **Bill's Sentiment Analysis:**

#### **Comparison Questions**

The IMR board should be summarizing the rationale as to their decisions objectively.

1. **How much subjectivity is seen in the IMR board's summaries?**
2. **Is there a difference in the subjectivity between those decisions that are upheld vs. overturned?**
3. **Is there a difference in the subjectivity in decisions by age range?**

#### **Brief Description of the Program**

This program (available in our Jupyter notebook) analyzes the unstructured text data. Ryan Summers cleansed this data and removed appropriate stop words to make the analysis more meaningful. The program then groups the data as needed to drill down to the analyses being performed.

Leveraging the following imports, the program performs sentiment analysis as it relates to subjectivity.

```
#Imports to support subjectivity analysis
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from textblob import TextBlob|
```

The raw numerical results are returned from 0-1 with 0 being completely objective, and 1 being completely subjective. The program categorizes the raw data into 5 buckets by

Team:

Ryan Summers, Bill Steel, Andrew Alford

---

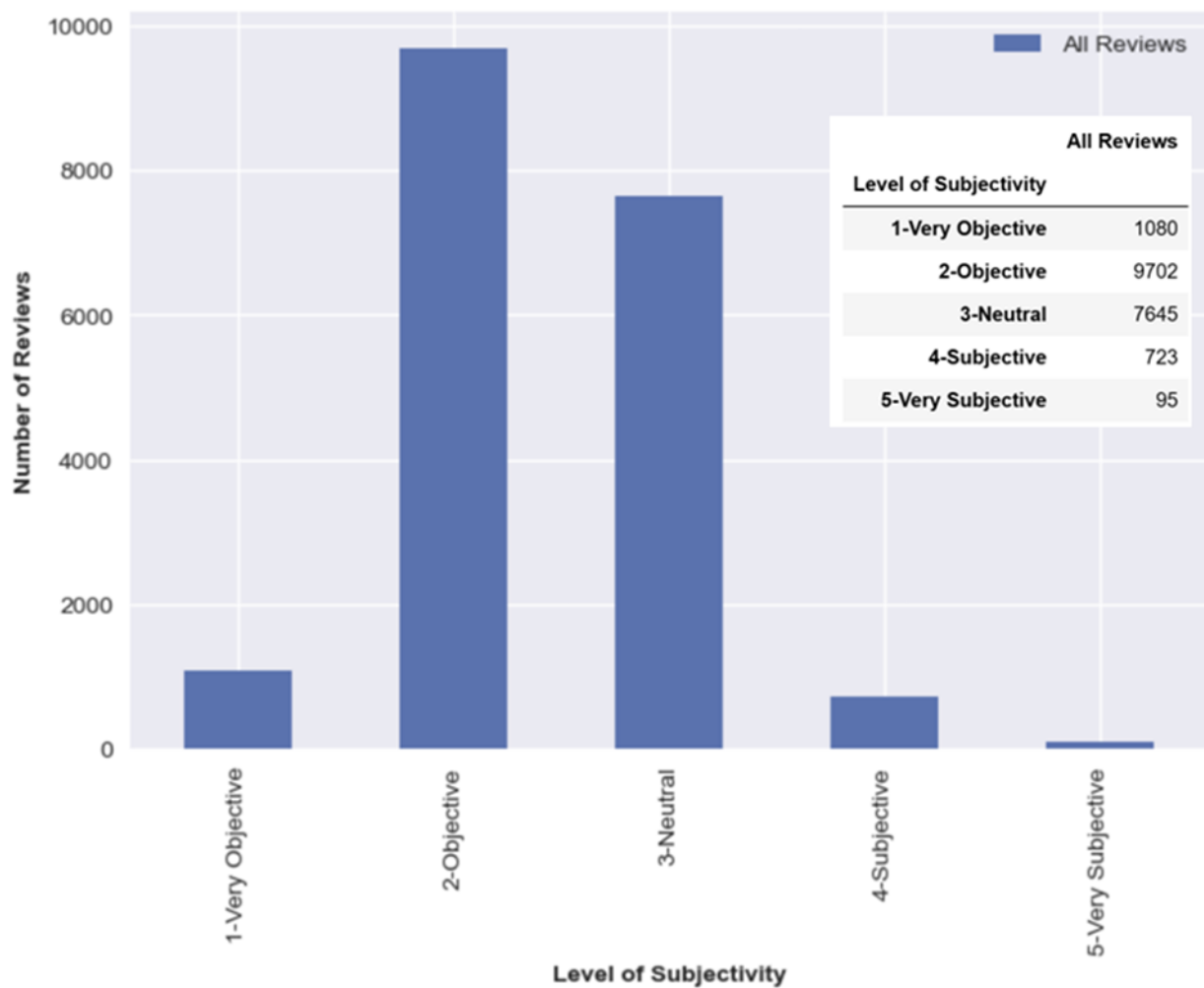
Category	Value range
1. Very Objective	0 to less than .2
2. Objective	.2 to less than .4
3. Neutral	.4 to less than .6
4. Subjective	.6 to less than .8
5. Very Subjective	.8 and greater

Following the analysis, the program generates the appropriate visualizations.

## **Output and Analysis**

*Question 1: How much subjectivity is seen in the IMR board's summaries?*

Based on the subjectivity analysis across all 19245 reviews, it's quite clear that the overwhelming majority are either objective/neutral. Well over 50% of the reviews are Objective/Very Objective with a small minority (<5%) being Subjective/Very Subjective. This certainly is encouraging as subjectivity shouldn't be a part of the review process. That said, subsequent analysis against the 818 subjective/very subjective reviews would be interesting.

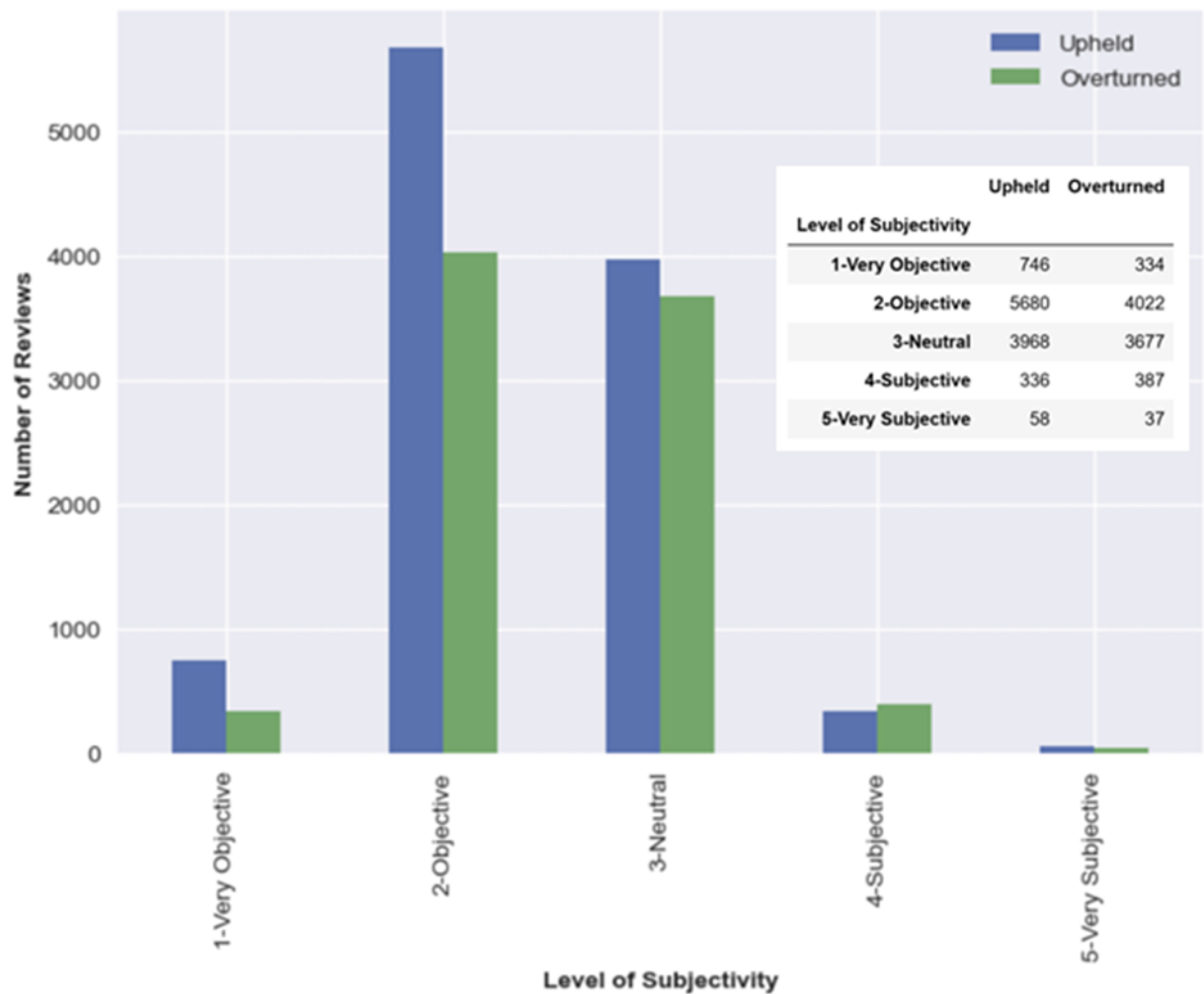


Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

*Question 2: Is there a difference in the subjectivity between those decisions that are upheld vs. overturned?*

Looking at the bar chart below, the IMR board's decisions tend to be more objective when the decision is to uphold the insurance company's denial of service.

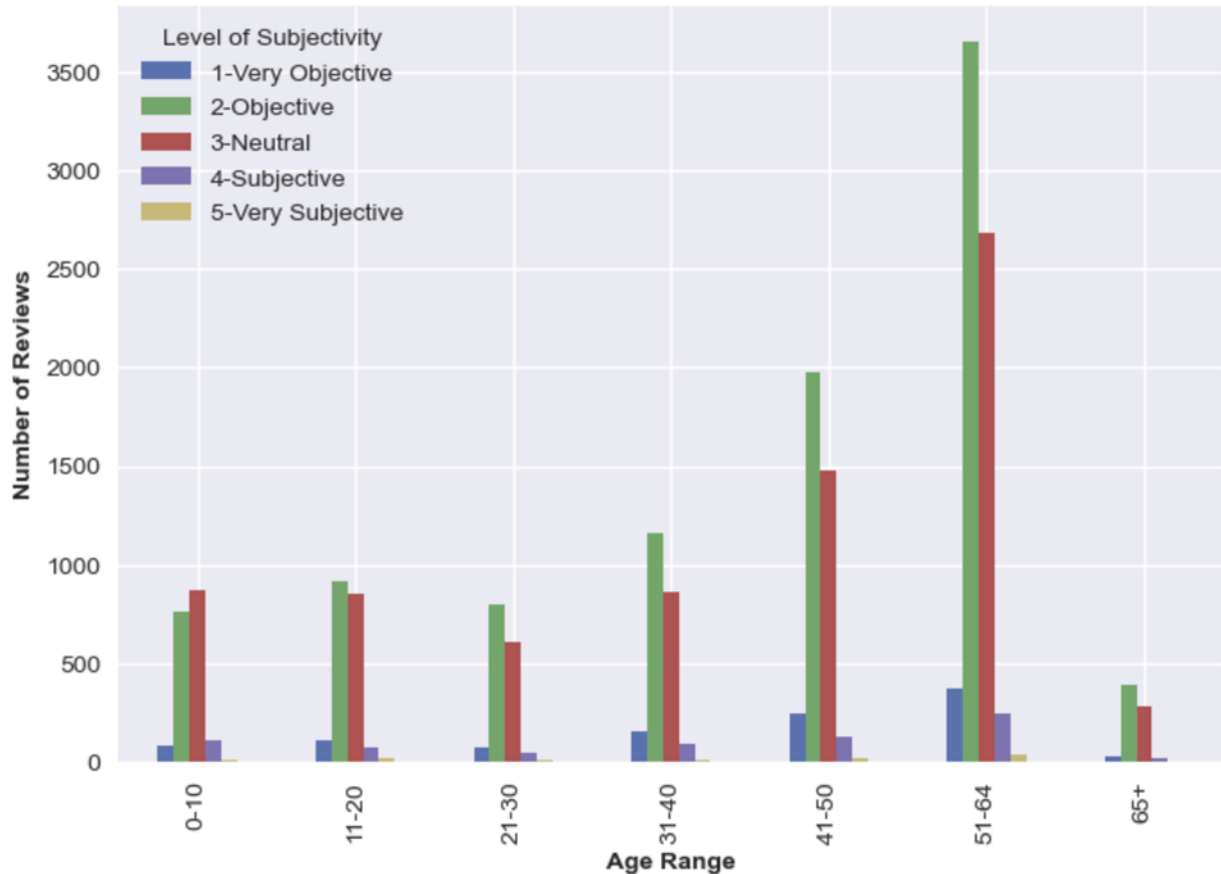


Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

*Question 3: Is there a difference in the subjectivity in decisions by age range?*

Looking at the bar chart below, the IMR board's summaries tend to follow the same profile as the overall analysis with regards to subjectivity by age group. Regardless of age group, the overwhelming majority are either objective/neutral. Objective/very objective results are over 50% of the total for all age groups with the exception being the 0-10 year old age group where slightly less than 50% of the reviews are Objective/Very Objective.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Sentiment Analysis:**

Naturally, the nature of discussing illness is an inherently emotionally difficult and uncomfortable subject. This gave our group the idea that it would be interesting and beneficial for health providers to have data on how their text summaries of each case are ranked on a positive/negative sentiment analysis. Similar to Bill's analysis on objectivity/subjectivity, these explanations are free form text consisting of around 300 words per review, the equivalent of about 18 novels at 1000 pages each.

### **Andrew's Sentiment Analysis:**

#### **Comparison Questions**

The IMR board should be summarizing the rationale as to their decisions neutrally.

1. **How neutral is the language in the summary column?**
2. **What is the most positive and negative sentiment among the top five diagnosis categories?**
3. **Using a word cloud visualization, what are the most common words seen in the summary column?**

### **Brief Description of the Program**

This program, which can be analyzed within our Jupyter Notebook, performs a sentiment analysis on the findings\_cleaned column of the data set. This column was cleansed by Ryan Summers to remove the appropriate stop words to improve the clarity of the words.

The exact line of code that is used to perform the sentiment analysis is as follows:

```
#Adding sentiment analysis and sentiment columns
from textblob import TextBlob

andrew_imr['sentiment'] = andrew_imr['findings_cleaned'].apply(lambda x:
TextBlob(x).sentiment.polarity)

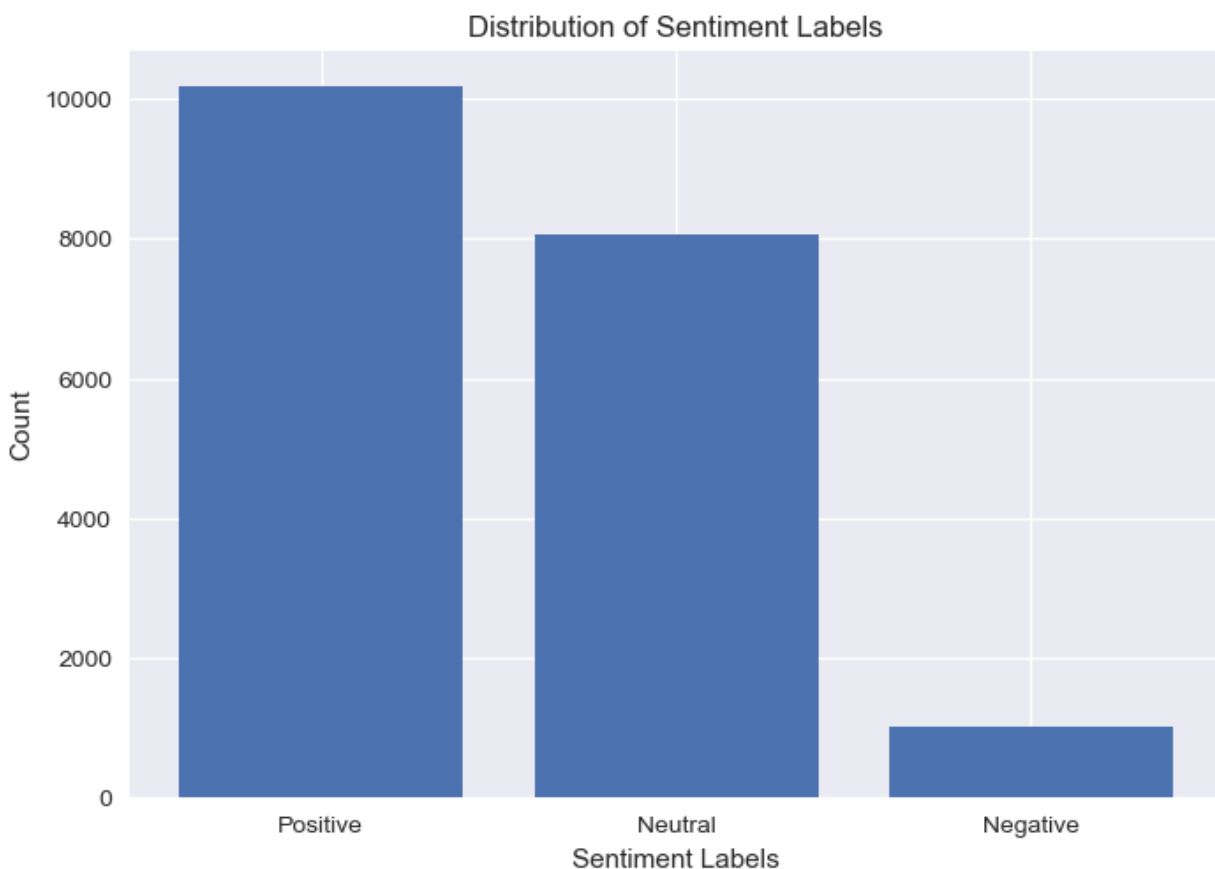
andrew_imr['sentiment_label'] = andrew_imr['sentiment'].apply(lambda x:
'Positive' if x > 0.05 else ('Negative' if x < -0.05 else 'Neutral'))
```

Each word is ranked on a scale from -1 to 1, with more negative words being considered negative numerically and positive sentiment words being numerically positive and neutral words being zero. I adjusted the values so that a word is considered neutral if it is within 5% of exactly zero.

## **Output and Analysis**

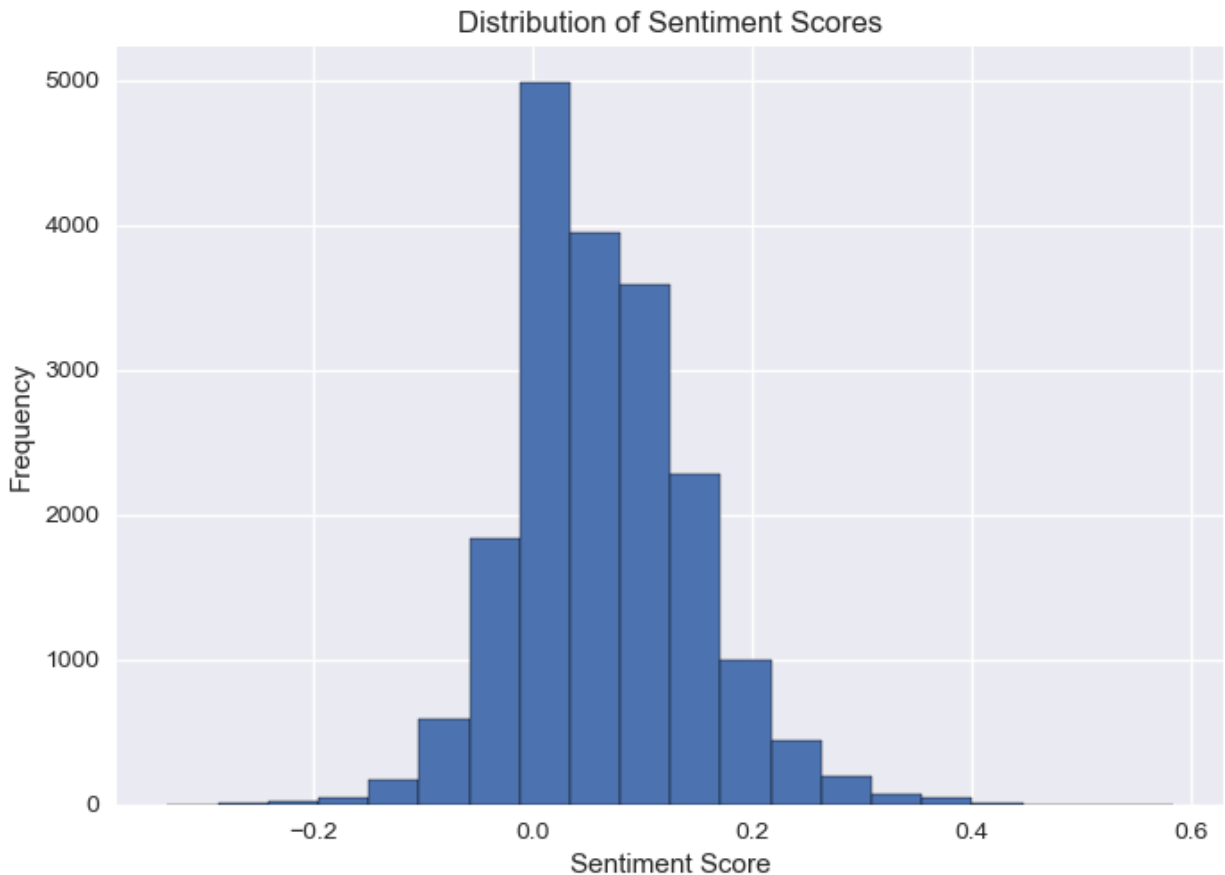
*Question 1: How neutral is the language in the summary column?*

Based on tens of thousands of words, there seems to be a strong mix of positive and neutral sentiments presented in the analysis. The positively ranked words ranked at over ten thousand, with the neutral words following closely behind at eight thousand. The negatively ranked words were strongly outnumbered, presenting roughly around one thousand negatively ranked words in total. It's also important to note that, within the second visualization presented here, the distribution of the words is quite small. With 0 being the middle, the overwhelming majority of words are within 0 and 0.4, with absolutely no words being ranked higher than 0.6. This seems to present the summary column in a highly neutral light, leaning to a slightly positive trend in sentiment.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---



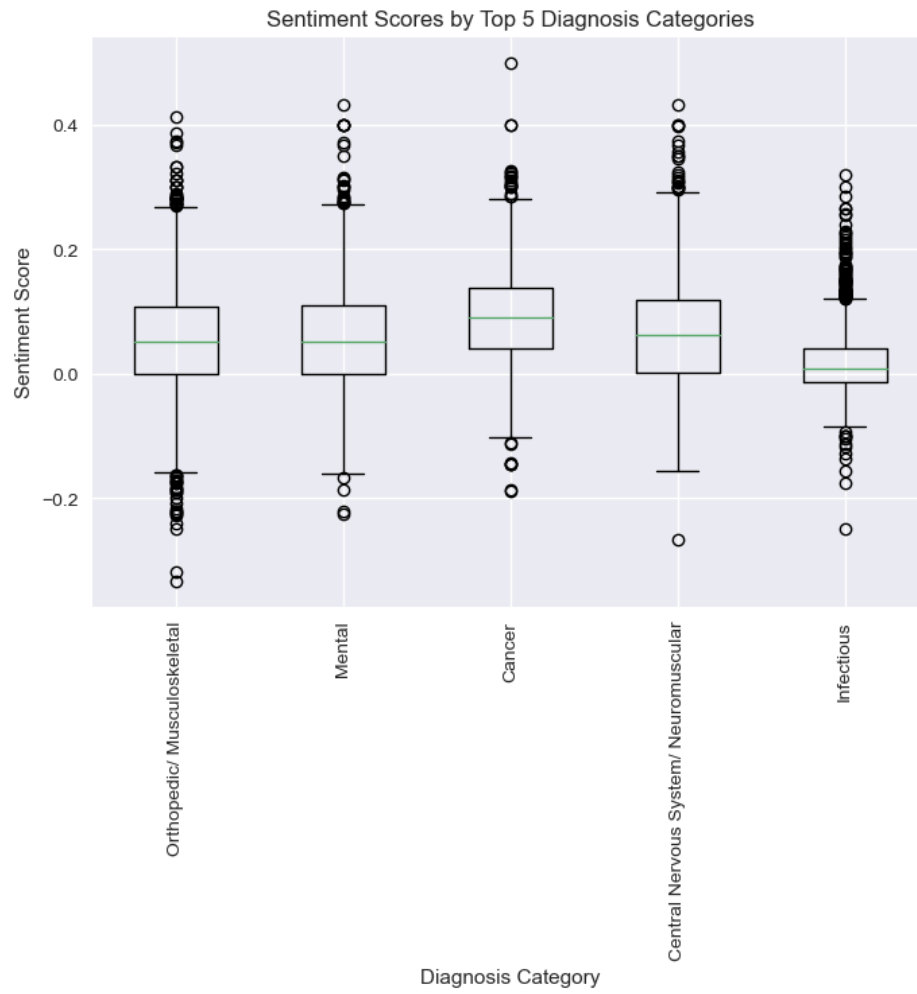
*Question 2: What is the most positive and negative sentiment among the top five diagnosis categories?*

Naturally, different diagnoses come with a variation in emotional responses to them, and perhaps a difference in how healthcare providers report them. Taking the top five most common diagnosis categories, the box plot below shows how each different diagnosis is ranked in terms of its positive/negative sentiment. The diagnosis that was most positively received was cancer, with an average of around 0.1. The most negatively ranked diagnosis was infectious, hovering just above 0.0, an almost perfectly neutral reaction. As you can see the deviation from one diagnosis to another is very small, pointing to the summary column being extremely neutral in sentiment.



Team:  
Ryan Summers, Bill Steel, Andrew Alford

---



Overall, there were four main conclusions that were garnered from the positive/negative sentiment analysis. The first main conclusion is that the findings\_cleaned column was largely neutral in nature, with a slight positive tone to it. The overwhelming majority of words ranked between 0.0 and 0.2, demonstrating just how truly neutral the wording was. The second main conclusion is that the five largest diagnosis categories were not all received the same sentiment wise. Cancer ranked as the more positive scoring at about 0.1 and the Infectious disease diagnosis category ranking as the lowest at just above 0.0. Lastly, I created a word cloud that visualized the most prevalent words within the column, noting that the removal of many unnecessary words was important for creating a valuable visualization. Some of the most prevalent words included year old, old female, evaluation, appropriate, and use.

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

### **Final Conclusions:**

A subsequent analysis examining the relationship between case determination and gender revealed statistically insignificant results, indicating no association between the decision-making process and the gender of the patient. This suggests that the determinations were primarily influenced by the unique characteristics of each individual case rather than any biases related to gender. Our findings emphasize that the decision-making process in these cases was driven by objective factors rather than influenced by gender.

In addition to the gender analysis, it is important to note that women accounted for 37.3% more cases than men in the dataset. Even when accounting for obstetrics and gynecology (OBGYN) cases, which are specific to women, there was still a 27.5% higher proportion of women's cases. It is crucial to emphasize that the cases reviewed were health plan denials, and while no biases were noted in the reviewers' determinations, the higher rate of cases where a women's health plan denies coverage indicates a notable trend. These findings shed light on potential disparities in health plan coverage for women, highlighting the need for further investigation into the reasons behind these discrepancies.

The age group with the most reviews was 51 to 64 years old, and the 65+ age group had the lowest number of IMR reviews. Challenges to service denial for children between the ages of 0-10 were successful over 67% of the time. Patients in the 65+ age group have a significantly low rate of reviews (and the lowest rate of successful reviews at about 37%). The low rate of reviews can be explained by the fact that Medicare is the typical insurance provider for Seniors and is not involved in the Independent Medical Review process that focuses on private insurance companies.

During subsequent analysis, Autism, Pediatric, and Infectious had significantly higher overturn rates than the other diagnosis categories. After a deeper analysis, Autism and Pediatric IMR overturned reviews were significantly higher in 2011 thus contributing significantly to the overall results. Infectious IMR overturned reviews were astronomically higher in 2015/2016 driving the overall results.

Orthopedic/Musculoskeletal, Mental, Central Nervous System/Neuromuscular, and Cancer were the most frequent diagnostic categories. Pharmaceutical treatments were the most common recommendations, with nearly 25% of the entire dataset. The determinations to overturn or uphold health plan decisions remained consistent from 2001 to 2014, but there was a sharp increase in cases in 2015 with overturned decisions outweighing upheld decisions for the first time, which was more pronounced in 2016.

Within the cancer portion of the project, there were a number of interesting insights. The first interesting insight was that recorded cancer rates fell steeply once patients reached the age group 65+, concluding that either a sizable portion died or stopped being recorded due to enrolling in medicare. Secondly, rates of cancer were generally equal between men and

Team:  
Ryan Summers, Bill Steel, Andrew Alford

---

women, with women being a slightly greater portion of the rates. The most prevalent form of cancer was breast cancer followed closely by prostate cancer. Lastly, PET scans and chemotherapy were the most common forms of cancer treatment.

**Sentiment Analysis final conclusions:**

Analyzing the written summaries from the IMR board shows that their reviews are overwhelmingly objective or neutral with less than 5% being subjective/very subjective. There does tend to be more objectivity in the summaries when the IMR board upholds an insurance company's decision to deny services as opposed to overturning the decision. That said, over 800 reviews out of 19,245 were subjective/very subjective. While indeed less than 5%, further analysis is encouraged as these types of medical reviews should never be subjective.

Within the positive/negative sentiment analysis, it clearly shows that the language used in the findings\_cleaned column were overwhelmingly neutral in nature with a slightly positive lean towards them and hardly any negativity. However, to juxtapose this there was essentially no language that was extremely positive, with the highest rating given to any word in the entire column being 0.5, only halfway to the most positive score possible. Secondly, an interesting conclusion is that Cancer received the most positive sentiment score at a little over 0.1 and Infectious disease ranked the lowest at almost perfectly 0.0, which clearly shows how little differentiation there was between the different diagnosis categories.

The analysis of the most common bigrams for men and women in IMR decisions revealed interesting patterns in regards to the type of medical conditions reviewed. Hepatitis C and back pain were common issues for both men and women, indicating their significance in the review process. Additionally, the presence of gender-specific conditions like prostate cancer for men and breast cancer for women reflects the consideration of specific health challenges unique to each gender.

These findings highlight the specific health concerns that are often addressed in IMR decisions for each gender. However, it is important to note that while certain medical conditions appeared with higher frequency, no evidence suggests any potential gender biases in the reviews.