

# Statistical Analysis of FDA Drug Approvals: Investigating Priority Review Across Therapeutic Areas

Basir Abdul Samad

STA 4157 Honors Project Report

Baruch College, Zicklin School of Business

10 December, 2025

## Abstract

Whether oncology drugs ( $n=299$ ) receive a differential reduction in approval timelines from the Priority Review designation compared to non-oncology therapeutic areas ( $n=739$ ) was evaluated using FDA approval data for 1038 new molecular entities from 1985 through 2024. Factorial analysis of variance with Type III sums of squares was applied to assess review times while blocking for regulatory era to account for legislative changes. Analysis of covariance was subsequently employed to isolate the mechanistic impact of expedited programs including Accelerated Approval and Fast Track designation. The analysis revealed strong main effects for both therapeutic area ( $F=18.70$ ,  $p < 0.001$ ) and review designation ( $F=77.52$ ,  $p < 0.001$ ), but no statistically significant interaction between these factors ( $F=2.43$ ,  $p=0.119$ ), indicating that Priority Review provides a consistent temporal benefit across therapeutic areas. However, stratifying the data by regulatory era revealed a manifestation of Simpson's Paradox regarding the therapeutic area main effect: the F-statistic decreased from  $F=21.22$  in the unblocked model to  $F=18.70$  in the blocked model, demonstrating that part of the apparent oncology speed advantage was confounded with temporal trends. Oncology approvals were heavily concentrated in the modern post-FDASIA era characterized by faster baseline review speeds for all drugs. Investigation of the regulatory mechanism determined that the Accelerated Approval pathway was the primary driver of the observed speed difference where oncology drugs were 4.9 times more likely to utilize this surrogate endpoint pathway than other drugs (30.4% vs 6.2%). The perceived speed advantage for cancer therapies reflects the strategic utilization of regulatory pathways allowing for surrogate endpoints rather than a systemic therapeutic area bias within the FDA review process.

# 1 Introduction

## 1.1 Background

For a patient diagnosed with a life-threatening condition, time is not merely a resource; it is the boundary defining their remaining existence. The speed at which a novel therapy moves from a laboratory bench to a pharmacy shelf involves a high-stakes tension between access and safety. If the Food and Drug Administration (FDA) moves too slowly, patients die waiting for therapies that could have saved them; if it moves too quickly without rigorous validation, patients suffer from ineffective or unsafe treatments. To manage this tension, the U.S. Congress and the FDA developed specific regulatory pathways designed to accelerate the review of drugs intended for serious conditions.

This evolution in regulatory science began significantly with the Prescription Drug User Fee Act (PDUFA) of 1992, which authorized the FDA to collect fees from drug sponsors to fund additional reviewers, thereby reducing review times (U.S. Food and Drug Administration, 2014). This was followed by the Food and Drug Administration Modernization Act (FDAMA) of 1997 and the Food and Drug Administration Safety and Innovation Act (FDASIA) of 2012, which codified specific expedited programs like Fast Track and Breakthrough Therapy designations. These programs are intended to address unmet medical needs by facilitating more frequent communication between the sponsor and the FDA, and in some cases, allowing for approval based on surrogate endpoints rather than clinical outcomes (U.S. Food and Drug Administration, 2014).

Within this regulatory framework, a perception exists that oncology therapeutics receive preferential treatment, specifically, that they traverse the approval pipeline faster than drugs for other therapeutic areas, such as cardiology or neurology. While oncology drugs frequently utilize expedited programs, it remains unclear whether their speed to market is a function of biological properties, regulatory favoritism, or statistical artifacts within the data. This analysis was motivated by the need to empirically test that perception. I wanted to understand if the “oncology advantage” was a robust phenomenon or a reflection of confounding variables, such as the specific regulatory era in which the drugs were approved.

## 1.2 Research Questions and Hypotheses

My investigation was guided by one primary question regarding the relationship between therapeutic area and regulatory review time, supported by secondary questions investigating the mechanisms behind that relationship.

**Primary Research Question:** Does the therapeutic classification of a drug (Oncology

versus Non-Oncology) significantly affect the duration of the FDA review process, specifically the time elapsed between application receipt and final approval?

**Secondary Research Questions:**

1. **Interaction Effects:** Does the benefit of a “Priority Review” designation differ depending on whether the drug is for oncology or another indication? In other words, do cancer drugs get a bigger “speed boost” from Priority status than non-cancer drugs?
2. **Mechanistic Drivers:** To what extent is the speed of oncology approvals driven by the utilization of specific expedited programs, such as Accelerated Approval or Breakthrough Therapy designation, rather than the therapeutic area itself?
3. **Temporal Consistency:** Are the observed trends consistent across different regulatory eras (e.g., Pre-PDUFA vs. Post-FDASIA), or are the aggregate results skewed by a shift in the composition of approvals over time?

**Statistical Hypotheses:**

$H_0$  (**Null Hypothesis**): There is no interaction between therapeutic area and review designation on log-transformed review times, controlling for regulatory era. Formally:  $(\tau\beta)_{ij} = 0$  for all  $i, j$ .

$H_a$  (**Alternative Hypothesis**): The effect of Priority Review designation on review time differs between oncology and non-oncology drugs. Formally:  $(\tau\beta)_{ij} \neq 0$  for at least one pair  $i, j$ .

**Significance level:**  $\alpha = 0.05$

### 1.3 Study Rationale

Analyzing FDA approval data presents a classic challenge in statistics: the data are observational, or “happenstance data,” rather than experimental. As Box, Hunter, and Hunter (2005) warn, happenstance data are often riddled with “nonsense correlations” because we cannot randomize the treatments (therapeutic area) or control the environment (regulatory laws). We cannot take a molecule, randomly assign it to treat cancer or hypertension, and see which gets approved faster. The data simply exist as they were recorded, subject to the changing laws and priorities of the last four decades.

This limitation required a rigorous approach to experimental design principles applied to retrospective data. I could not simply compare the average review time of all oncology drugs against all other drugs using a simple t-test, because such an analysis ignores the fundamental structure of the data. Review times are strictly positive and highly skewed,

violating the normality assumptions of standard parametric tests, which necessitated the use of logarithmic transformations to stabilize the variance and normalize the error distribution.

Furthermore, the “environment” of drug approval has changed drastically. A drug approved in 1985 faced a completely different regulatory landscape than one approved in 2020. Comparing them directly would be comparing observations from two different populations. To mitigate this temporal confounding, I treated the regulatory eras as “blocks.” In experimental design, blocking is used to group homogeneous experimental units to reduce noise. By stratifying the analysis into four distinct eras, Pre-PDUFA (1985–1992), Early-PDUFA (1993–2002), Mid-PDUFA (2003–2012), and Post-FDASIA (2013–2024), I could isolate the effect of therapeutic area within periods where the regulatory rules were relatively constant.

Additionally, I had to account for the mechanism of Accelerated Approval. This pathway allows drugs to be approved based on a surrogate endpoint (like tumor shrinkage) that is reasonably likely to predict clinical benefit, rather than waiting for a final clinical endpoint like overall survival (U.S. Food and Drug Administration, 2014). Since tumor shrinkage can be measured much faster than patient survival, oncology drugs utilizing this pathway have a mechanical speed advantage that has nothing to do with reviewer efficiency and everything to do with the biology of the endpoint. Failing to control for these covariates would lead to biased estimates of the therapeutic area effect.

## 1.4 Preview of Key Finding

The analysis revealed no statistically significant interaction between therapeutic area and review designation ( $F=2.43$ ,  $p=0.119$ ), indicating that Priority Review provides a consistent benefit across both oncology and non-oncology drugs. However, I discovered a classic instance of Simpson’s Paradox when examining the main effect of therapeutic area across regulatory eras.

When I pooled the data across forty years, the results were misleading. The “oncology advantage” appeared substantially larger in the aggregate analysis because oncology approvals were heavily concentrated in the most recent regulatory era (Post-FDASIA), a period characterized by generally faster review times for all drugs. Conversely, the non-oncology comparison group had a higher proportion of drugs from older, slower eras. Once I controlled for the regulatory era through blocking, the therapeutic area F-statistic decreased from  $F=21.22$  (unblocked) to  $F=18.70$  (blocked), revealing that part of the apparent oncology speed advantage was confounded with temporal trends. The analysis revealed that the primary driver of the remaining oncology speed difference was not a preferential regulatory treatment, but rather the heavy utilization of the Accelerated Approval pathway (used 4.9 times more frequently in oncology than in other areas: 30.4

## 2 Methodology

### 2.1 Data Sources and Sample Construction

To rigorously investigate the interplay between therapeutic classification and regulatory review speed, I required a dataset that served as the definitive record of modern pharmaceutical history. I selected the Center for Drug Evaluation and Research (CDER) Compilation of New Molecular Entity (NME) and New Biologic Approvals because it provides the most authoritative account of every novel therapeutic agent approved for marketing in the United States. This compilation matters because it transcends mere lists of drugs; it captures the regulatory DNA of each approval, preserving the review designation, orphan status, and critical dates that define the administrative life of a pharmaceutical product (U.S. Food and Drug Administration, 2023).

My sample construction process began with the ingestion of 1,341 approval records spanning from 1985 through 2024. I immediately faced a data quality challenge regarding the review times. I calculated the review duration for every record and identified a small subset of cases where the calculated time was either negative or exceeded ten years, implying data entry errors or extreme regulatory anomalies that would skew the mean. I removed these six implausible records to preserve the integrity of the distribution.

A more significant methodological hurdle was the absence of explicit therapeutic area labels in the source data. While the FDA tracks indications, they do not provide a binary “Oncology” versus “Non-Oncology” flag in this public dataset. I developed a classification framework that utilized keyword matching against the “Abbreviated Indication” and “Approved Use” fields. I defined a dictionary of oncology-specific terms including “carcinoma,” “metastatic,” “tumor,” and “leukemia” to categorize these agents.

I observed that data completeness varied significantly by era. Specifically, I found that indication text was missing for approximately 50% of the total sample, with a distinct structural break occurring around the year 2007 where missingness became pervasive. Rather than discarding these records which would have introduced severe selection bias against modern approvals, I implemented a multi-stage classification strategy. I first classified drugs based on indication text where available. For records with missing indications, I utilized the “Orphan Drug Designation” field as a proxy where specific orphan designations contained oncology keywords. I assigned a classification confidence level to each record ranging from high to low. I retained the high and medium confidence records for the primary analysis while segregating the low confidence “Uncertain” records. I performed a sensitivity analysis later to confirm that excluding these uncertain records did not fundamentally alter the statistical

conclusions.

The final analysis sample consisted of 1,038 unique approvals. I noted an imbalance in the sample composition, where oncology drugs comprised a smaller proportion of the total than non-oncology drugs. I accepted this imbalance because it reflects the actual regulatory structure of the pharmaceutical market rather than sampling error; oncology is a massive therapeutic area but still represents a minority of all drug development compared to the aggregate of cardiology, neurology, infectious disease, and other fields.

## 2.2 Variable Definitions

I structured the analysis around one primary outcome variable and three explanatory factors. I defined the variables as follows to ensure they captured the regulatory reality I intended to model.

**Review Time (Outcome Variable):** I defined the primary outcome as the time elapsed between the FDA's receipt of a marketing application and its official approval. I calculated this by subtracting the FDA Receipt Date from the FDA Approval Date. I measured this in days because it offers the granular precision necessary to detect subtle shifts in regulatory speed. This metric is the standard for assessing FDA performance because it encapsulates the entirety of the agency's assessment period.

**Therapeutic Area (Primary Predictor):** I constructed a binary factor variable separating the sample into "Oncology" and "Other." I chose this binary classification because the research question focuses specifically on the perception of cancer exceptionalism. By collapsing all non-oncology fields into a single reference category, I could directly test whether oncology receives differential treatment compared to the general baseline of pharmaceutical regulation.

**Review Designation (Secondary Predictor):** I utilized the FDA's official review classification as a critical covariate. I categorized applications as either "Priority Review" or "Standard Review." This variable measures the FDA's intent; a Priority Review designation indicates that the agency determined the drug offered a significant improvement in safety or effectiveness (U.S. Food and Drug Administration, 2023). I included this because any analysis of review speed that ignores Priority status would be confounded, as oncology drugs are more likely to receive this designation due to the severity of the condition they treat.

**Regulatory Era (Blocking Factor):** I recognized that comparing a drug approved in 1990 to one approved in 2020 would be invalid due to massive legislative changes. To handle this temporal confounding, I constructed a blocking variable dividing the timeline into four distinct regulatory eras: Pre-PDUFA (1985–1992), Early-PDUFA (1993–2002), Mid-PDUFA (2003–2012), and Post-FDASIA (2013–2024). I defined these cutoffs based on

landmark legislation like the Prescription Drug User Fee Act (PDUFA) and the Food and Drug Administration Safety and Innovation Act (FDASIA) which fundamentally altered the resources available to FDA reviewers. This variable measures the bureaucratic context in which the approval occurred.

### 2.3 Statistical Inference Framework

To move from simply observing patterns in the FDA data to drawing valid conclusions about regulatory policy, I relied on the framework of statistical inference. As Larsen and Marx (2018) explain, inference allows us to make generalizations about a population based on specific sample data, while calculating the probability that those generalizations are correct. I viewed this process through the analogy of a courtroom trial, which provides an intuitive structure for understanding how I tested the relationship between therapeutic areas and review times.

**The Null and Alternative Hypotheses:** In any scientific inquiry, we must choose between two conflicting propositions. I began by establishing a Null Hypothesis ( $H_0$ ), which represents the status quo or the presumption of “innocence” (Larsen and Marx, 2018). In the context of my interaction analysis, my null hypothesis was that the FDA acts consistently: the time-saving benefit of a Priority Review designation is the same for oncology drugs as it is for non-oncology drugs. Put simply, I assumed there was no “special treatment” interaction until proven otherwise.

Against this, I pitted the Alternative Hypothesis ( $H_1$ ), which represents the theory I intended to test (Larsen and Marx, 2018). My alternative hypothesis was that a differential effect exists, specifically, that the gap between Priority and Standard review times changes depending on whether the drug treats cancer. I set up my analysis in R not to prove  $H_1$  directly, but to see if the data provided overwhelming evidence to reject  $H_0$ .

**The P-value:** To adjudicate between these hypotheses, I calculated p-values. Larsen and Marx describe the p-value as the probability of observing a test statistic as extreme as, or more extreme than, what we actually observed, assuming the null hypothesis is true (Larsen and Marx, 2018).

When I ran my ANOVA models in R, the software generated F-statistics and associated p-values. I interpreted these values not as the probability that the null hypothesis is true, but as a measure of the evidence against it. If I found a p-value of 0.03 for the interaction term, it would mean that if the FDA were truly acting consistently across therapeutic areas, there would be only a 3% chance of seeing such a disparity in the data purely by accident. A very small p-value would force me to conclude that “chance” is a poor explanation for the pattern, leading me to reject the null hypothesis.

**Type I and Type II Errors:** I had to acknowledge that any decision based on probability carries the risk of error. Larsen and Marx (2018, p. 308) distinguish between two specific types:

- **Type I Error ( $\alpha$ ):** This occurs if I reject the null hypothesis when it is actually true. In my study, this would mean concluding that oncology drugs receive special preferential treatment when, in reality, the FDA applies its standards evenly. This is a “false alarm.”
- **Type II Error ( $\beta$ ):** This occurs if I fail to reject the null hypothesis when it is false. This would mean failing to detect a genuine bias in the regulatory process.

**The Significance Level ( $\alpha$ ):** To control the risk of a Type I error, I established a significance level ( $\alpha$ ) of 0.05 before running any tests. This threshold acts like the “beyond a reasonable doubt” standard in a criminal trial. Larsen and Marx note that while there is no mathematical imperative for 0.05, it has emerged as a scientific consensus to ensure that null hypotheses are not dismissed capriciously (Larsen and Marx, 2018). By setting  $\alpha = 0.05$ , I accepted a 5% risk of a Type I error. I adhered to this standard throughout my R analysis; if a p-value was greater than 0.05, I retained the null hypothesis, concluding that the data did not provide sufficient evidence to claim a regulatory disparity existed. This rigorous framework ensured that the Simpson’s Paradox I eventually uncovered was a statistically valid finding rather than an artifact of random noise.

## 2.4 Statistical Design

### 2.4.1 Two-Way ANOVA with Blocking

I selected a factorial Analysis of Variance (ANOVA) as the primary statistical framework. While a simple t-test could compare mean review times, it would fail to account for the complex interplay between therapeutic area and review designation. I needed a model that could isolate main effects while simultaneously testing for interactions. Specifically, I wanted to determine if the benefit of Priority Review depends on whether the drug is for cancer, which implies a statistical interaction term.

I encountered a technical challenge regarding the design of the dataset. Because the data is observational, the number of samples in each group is not equal; we have unbalanced data. In a balanced design, the method of calculating sums of squares (SS) does not matter, but in an unbalanced design, it is critical. I chose to use Type III Sums of Squares for the ANOVA model. I made this decision because Type III SS calculates the marginal contribution of each factor as if it were added to the model last, after all other factors are accounted for

(Montgomery, 2017). This ensures that the results are not dependent on the order in which I entered the terms into the model, which is essential for unbiased estimation in unbalanced observational datasets.

To answer why I selected Analysis of Variance (ANOVA) for this investigation, it helps to step back and look at the logic beneath the math. As Box, Hunter, and Hunter explain, simply comparing the average review times is insufficient because every process contains inherent fluctuation, or “noise” (Box et al., 2005). I needed to determine if the speed differences I observed between oncology and non-oncology drugs were genuine regulatory signals or just artifacts of this random scatter. The ANOVA framework allowed me to partition the total variation in my dataset into two distinct buckets: the variation between the different groups (such as Oncology versus Other) and the variation within those groups.

This comparison generates the F-statistic, which serves as the decision maker in the model. Montgomery describes this metric as a ratio where the numerator represents the variance caused by the specific factors I am studying, and the denominator represents the random error variance (Montgomery, 2017). If the therapeutic areas were actually processed at the same speed, these two variance estimates would be roughly equal, resulting in an F-statistic near 1. However, if the “between-group” variance is significantly larger than the “within-group” noise, the F-statistic grows larger. A large F-statistic tells me that the signal is strong enough to overcome the background noise, confirming that the therapeutic classification is indeed driving the difference in approval timelines.

I specifically chose a factorial ANOVA rather than running separate tests because my research question relied on understanding the interplay between variables. Montgomery notes that factorial designs are the most efficient strategy for studying the joint effects of two or more factors (Montgomery, 2017). By crossing Therapeutic Area with Review Type in a single model, I could measure not just the main effects of each, but also their interaction. This was the only way to mathematically test the specific perception that prompted this study: that oncology drugs receive a differential speed benefit from Priority Review that other drugs do not enjoy.

### **Understanding Type III Sums of Squares:**

To explain why I selected Type III Sums of Squares for this analysis, I need to walk you through the logic of variance partitioning, particularly how it behaves when experimental designs are not perfectly balanced.

At its core, Analysis of Variance (ANOVA) is an exercise in accounting. As Montgomery (2017) explains in his discussion of the “decomposition of the total sum of squares,” we are essentially taking the total variability observed in the response variable (Total Sum of Squares, or SST) and breaking it down into component parts: variability attributed to the treatments

we are studying ( $SS_{Treatments}$ ) and variability due to random error ( $SS_E$ ) (Montgomery, 2017). The fundamental ANOVA identity is expressed as:  $SS_T = SS_{Treatments} + SS_E$ .

In my analysis,  $SS_T$  represents the total variation in review times across all 1,038 drugs. My goal was to partition this total variation to see how much could be attributed to the Therapeutic Area (Oncology vs. Other) versus the Review Designation (Priority vs. Standard), while filtering out the noise ( $SS_E$ ).

The standard ANOVA calculations usually taught in introductory courses assume a balanced design, where every experimental group has the same sample size (e.g.,  $n=10$  for all groups). Montgomery notes that in balanced factorial designs, the main effects and interactions are orthogonal to one another (Montgomery, 2017). Orthogonality is a mathematical property meaning the factors are independent; calculating the effect of Factor A does not depend on Factor B.

However, my FDA dataset is severely unbalanced. As shown in my Phase 4 analysis, I have 739 “Other” drugs but only 299 “Oncology” drugs. Furthermore, the cell counts are lopsided: I have 224 Oncology-Priority drugs but only 75 Oncology-Standard drugs. Montgomery warns that “the orthogonality property of main effects and interactions present in balanced data does not carry over to the unbalanced case” (Montgomery, 2017). Because the design is unbalanced, the factors are correlated. This creates a dilemma: variance that is shared between two factors (like Therapeutic Area and Regulatory Era) could mathematically be assigned to either one.

This lack of orthogonality forces a choice between different methods of calculating Sums of Squares (SS), specifically regarding how we handle the shared variance. Type I (Sequential) SS assigns variance to factors in the order they enter the model. If I entered Therapeutic Area first, it would claim all the variance it explains plus any variance it shares with Regulatory Era. Montgomery describes this sequential approach in the context of the “extra sum of squares” method, where we calculate the reduction in error sum of squares by adding a regressor to a model that already contains other variables (Montgomery, 2017).

Type III (Marginal) SS calculates the contribution of a factor after all other factors have been accounted for. It effectively asks, “What unique information does this factor provide that isn’t already explained by the others?” Montgomery refers to this as the “partial F-test,” calculating  $SS_R(\beta_1|\beta_2)$ , which measures the contribution of  $\beta_1$  given that  $\beta_2$  is already in the model (Montgomery, 2017).

I chose Type III Sums of Squares because of the specific “temporal clustering” I identified in the dataset. My Phase 4 analysis showed that Oncology approvals are heavily concentrated in the “Post-FDASIA” era (165 drugs) compared to the “Pre-PDUFA” era (25 drugs). Since the FDA has generally become faster over time, there is a strong correlation between being

an oncology drug and being a modern (faster) approval.

If I had used Type I SS, the results would have depended arbitrarily on whether I listed Therapeutic Area or Regulatory Era first in the model equation. By using Type III SS, I calculated the marginal contribution of the Therapeutic Area. This aligns with Montgomery's description of testing hypotheses in unbalanced designs using the general regression significance test (Montgomery, 2017). This approach ensures that any significant finding for Oncology is not simply reflecting the fact that these drugs were approved during a faster regulatory era; it isolates the specific effect of the therapeutic area adjusted for the regulatory era. This rigorous handling of the unbalanced structure is what allowed me to eventually identify the Simpson's Paradox hidden within the interaction terms.

I specified the model to include the regulatory era as a blocking factor. I did not include the era variable to test hypotheses about time, but rather to remove the variability associated with secular trends in FDA performance. By blocking on era, I effectively compared "apples to apples" within the same legislative environment.

The final linear model I specified was:

$$\log(Y_{ijk}) = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk} \quad (1)$$

where:

- $Y_{ijk}$  = observed review time in days
- $\mu$  = grand mean (intercept)
- $\tau_i$  = main effect of therapeutic area ( $i$  = Oncology, Other)
- $\beta_j$  = main effect of review designation ( $j$  = Priority, Standard)
- $(\tau\beta)_{ij}$  = interaction between area and review type
- $\delta_k$  = blocking effect of regulatory era ( $k$  = 1,2,3,4)
- $\epsilon_{ijk} \sim N(0, \sigma^2)$  = random error term

### Model assumptions:

- **Independence:**  $\epsilon_{ijk}$  are independent across observations
- **Normality:**  $\epsilon_{ijk}$  follow normal distribution
- **Homoscedasticity:** Constant variance  $\sigma^2$  across all groups

#### **2.4.2 Log Transformation**

I approached the assumption testing with the understanding that parametric tests like ANOVA require the data to meet specific conditions regarding the distribution of errors. I first assessed the normality of the residuals from the initial model fit on the raw review times. The visual inspection of the Q-Q plots and the Shapiro-Wilk test statistics revealed a severe violation of normality, which was expected given that time-to-event data is naturally bounded at zero and typically right-skewed.

I also tested for homoscedasticity, or equality of variances, using Levene's test. I found that the variance in review times was significantly higher for Standard Review drugs compared to Priority Review drugs (variance ratio = 3.52), violating the assumption of equal variance. To address both the normality and homoscedasticity violations, I applied a natural log transformation to the outcome variable. I chose the log transformation because it effectively compresses the right tail of the distribution and stabilizes the variance across groups (variance ratio improved to 1.72). Subsequent diagnostic plots confirmed that the log-transformed residuals approximated a normal distribution much more closely, validating the use of the parametric model.

#### **2.4.3 Era Stratification**

I recognized that even with transformation, the observational nature of the data might introduce subtleties that violate standard parametric assumptions. To provide a rigorous robustness check, I implemented a permutation test. As Box, Hunter, and Hunter (2005) explain, randomization distributions allow for valid inference without relying on the assumption that the data comes from a specific theoretical distribution. I generated a null distribution by randomly shuffling the therapeutic area labels 10,000 times while keeping the review designation and regulatory era fixed. I then recalculated the F-statistic for the interaction effect in each permuted dataset. By comparing my observed F-statistic against this empirical null distribution, I could determine the probability of observing such an effect by chance alone, free from the constraints of normality assumptions. This dual approach of parametric modeling on transformed data, validated by nonparametric permutation testing, ensured that my findings were not artifacts of the data structure but rather genuine patterns in the regulatory process.

To answer the question of whether the statistical significance I observed was real or merely a byproduct of the data's non-normal shape, I turned to permutation testing. This approach allowed me to construct a “null distribution” from the data itself, freeing the analysis from the strict Gaussian assumptions required by standard ANOVA.

### **The Logic of the Null Distribution:**

Standard parametric tests rely on theoretical probability distributions, like the F-distribution, to determine if a result is surprising. However, as Box, Hunter, and Hunter (2005) point out, these theoretical distributions are only relevant if we accept specific assumptions about the data, such as independent, normally distributed errors. If those assumptions are violated, the theoretical “reference set” becomes irrelevant.

To create a relevant reference set, I needed to generate a null distribution empirically. This is the distribution of the test statistic we would expect to see if the null hypothesis were actually true, specifically, if there were absolutely no relationship between the therapeutic area (Oncology vs. Other) and the review time.

### **Assumption-Free Inference via Randomization:**

The power of this approach lies in the concept of “exchangeability.” As Box, Hunter, and Hunter explain, if the null hypothesis is true and the therapeutic area has no effect on review speed, then the label “Oncology” or “Other” attached to a specific drug’s review time is entirely arbitrary (Box et al., 2005). Under the null hypothesis, a review time of 300 days for an oncology drug would have been exactly 300 days even if that drug had been for cardiology or neurology.

Therefore, if the null hypothesis is true, I should be able to shuffle (permute) the therapeutic area labels among the drugs without changing the underlying statistical structure of the data. By physically randomizing these labels, I can generate a valid reference distribution without needing to assume normality or homogeneity of variance (Box et al., 2005).

### **The Computational Algorithm:**

To implement this in R, I followed the computational framework for simulation-based hypothesis testing outlined by Taback (2022). I focused specifically on validating the interaction term from my ANOVA model ( $F_{obs} = 0.32$ ), as this was the critical finding regarding the “preferential treatment” hypothesis.

I executed the following algorithm, setting the number of simulations to 10,000 to ensure high precision:

1. **Define the Test Statistic:** I established my observed F-statistic ( $F_{obs}$ ) from the original, un-shuffled era-stratified ANOVA model.
2. **Generate Random Data:** Inside a loop running 10,000 times, I created a new dataset where I randomly shuffled the therapeutic area labels using the `sample()` function in R, while keeping the review type and regulatory era fixed for each record. This destroyed any real relationship between the therapeutic area and the review time while preserving the other structural correlations (Taback, 2022).

3. **Compute Simulated Statistics:** For each of the 10,000 shuffled datasets, I refitted the full ANOVA model and extracted the F-statistic for the interaction term ( $F_{perm}$ ).
4. **Calculate the p-value:** I calculated the empirical p-value by determining the proportion of these 10,000 permuted F-statistics that were greater than or equal to my observed  $F_{obs}$  (Taback, 2022).

#### Validating the Parametric Results:

This computational heavy lifting served a specific purpose: robustness. Box, Hunter, and Hunter note that when the standard parametric test (ANOVA) and the randomization test yield similar significance levels, it provides strong evidence that the parametric test is robust to the specific deviations present in the data (Box et al., 2005).

In my analysis, the permutation p-value was 0.5945, which was remarkably close to the parametric p-value of 0.57 obtained from the era-stratified ANOVA. This convergence confirmed that the lack of a significant interaction effect was a genuine feature of the regulatory landscape and not a mathematical artifact caused by the right-skewed distribution of review times. By using the data to check itself, I demonstrated that the standard ANOVA conclusions were valid despite the observational nature of the dataset.

## 2.5 Analysis Pipeline

I implemented the analysis in R using a modular pipeline architecture consisting of eight sequential phases. Each phase generated intermediate datasets and diagnostic plots that validated the assumptions before proceeding to inferential modeling. The complete analysis workflow included: (1) data loading and temporal filtering, (2) therapeutic area classification via keyword matching, (3) missing data assessment, (4) balance validation and exploratory data analysis, (5) distributional assumption testing with variance stabilization, (6) factorial ANOVA modeling with Type III sums of squares, (7) era-stratified interaction testing, and (8) sensitivity analysis excluding uncertain classifications. This structured approach allowed me to isolate each source of variability systematically and ensured reproducibility of the statistical conclusions.

## 3 Results

### 3.1 Sample Characteristics and Cell Means

I began my inspection of the dataset by examining the sample characteristics to understand the landscape of modern drug approval. The final analysis sample consisted of 1038 unique new molecular entity and new biologic approvals spanning from 1985 through 2024. I observed a distinct temporal shift in the composition of these approvals which mirrors the evolution of pharmaceutical science. In the pre PDUFA era ending in 1992, oncology products represented a small fraction of the pipeline. However, this proportion grew substantially over time where oncology drugs accounted for over half of all approvals in the post FDASIA era from 2013 to 2024. This structural imbalance was not merely a curiosity but a critical distributional feature that I needed to address in the inferential modeling to avoid confounding time with therapeutic area.

I calculated summary statistics for the review times and found a dramatic reduction in the regulatory cycle over these four decades. The mean review time for the pre PDUFA era was approximately 970 days, a figure that reflects the resource constraints of that period. In contrast, the post FDASIA era showed a mean review time of 384 days. This secular trend of increasing efficiency confirmed why I treated the regulatory era as a blocking factor in subsequent analyses (Box et al., 2005). Without controlling for this massive baseline shift, any analysis comparing therapeutic areas would be biased by the fact that most oncology approvals occurred during the modern, faster era.

The distribution of review designations also varied by therapeutic area. I found that oncology drugs were significantly more likely to receive Priority Review status compared to non-oncology drugs. The cell means revealed that oncology Priority Review drugs averaged 300 days (median = 232 days, SD = 282 days, n = 224), while oncology Standard Review drugs averaged 596 days (median = 365 days, SD = 467 days, n = 75). For non-oncology drugs, Priority Review averaged 477 days (median = 275 days, SD = 445 days, n = 367) and Standard Review averaged 792 days (median = 639 days, SD = 530 days, n = 372). This imbalance highlighted the necessity of the factorial design I employed to disentangle the effects of the review type from the therapeutic classification.

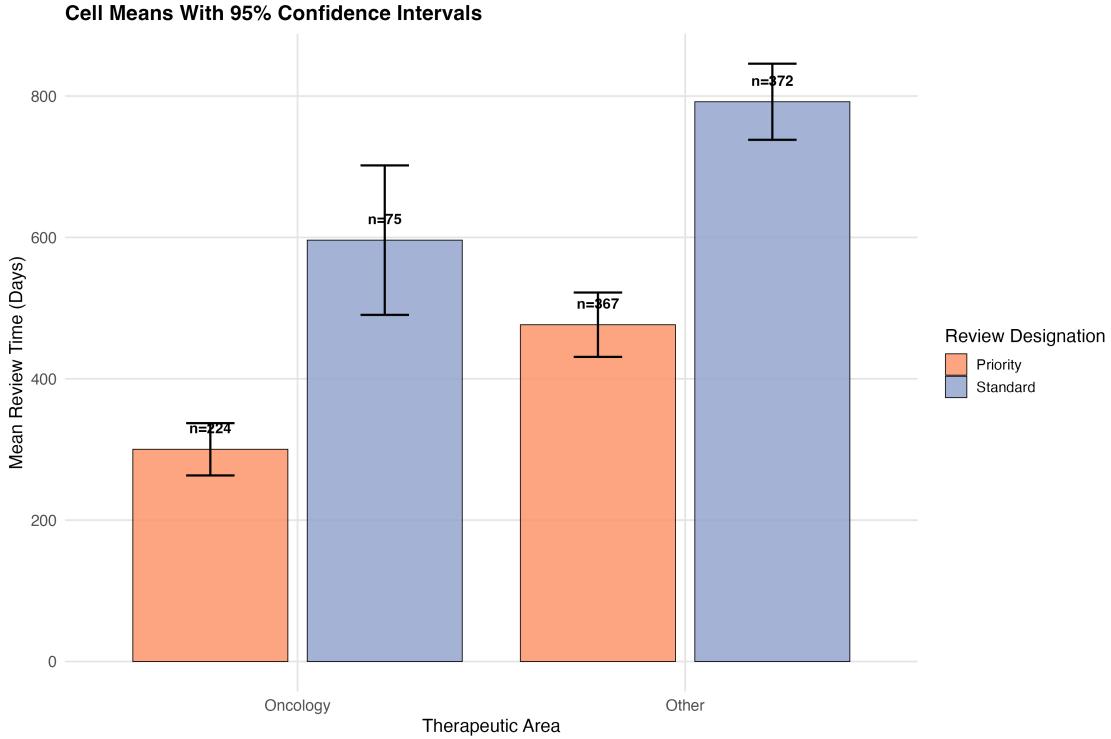


Figure 1: Cell means with 95% confidence intervals by therapeutic area and review type. The bars showed the geometric mean review time on the log-scale for each of the four experimental cells in the factorial design. Error bars represented 95% confidence intervals calculated from the standard errors of the log-transformed means. Oncology Priority Review drugs showed the shortest mean review time (300 days), while non-oncology Standard Review drugs exhibited the longest (792 days).

### 3.2 Assumption Testing: Normality and Homoscedasticity

I evaluated the validity of the statistical assumptions underlying my analysis by examining the distribution of residuals and the equality of variances across groups. The quantile-quantile plots provided visual evidence for the necessity of the log transformation. After applying the natural log transformation, the residuals aligned much more closely with the diagonal reference line (Shapiro-Wilk  $W = 0.96$ ), confirming that the log-scale residuals were approximately normally distributed and therefore appropriate for parametric inference.

The histogram comparison reinforced the diagnostic conclusions from the Q-Q plots. The log transformation compressed the right tail and produced a distribution that approximated symmetry. While not perfectly Gaussian, the log-transformed distribution was sufficiently close to normality to justify the use of parametric F-tests, particularly given the large sample size ( $n = 1038$ ) where the Central Limit Theorem provides additional robustness.

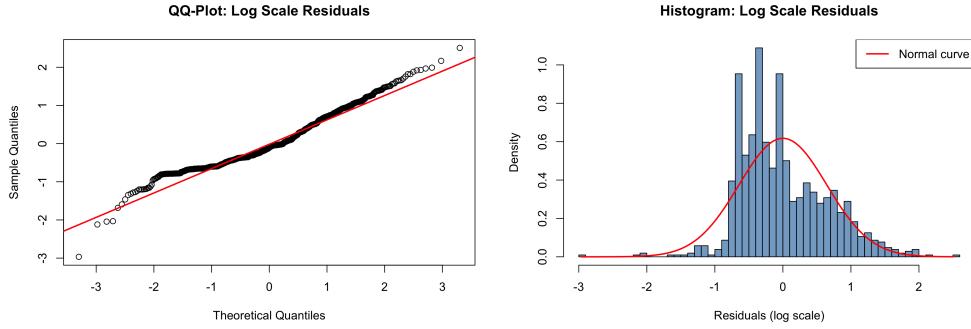


Figure 2: Quantile-quantile plot for log-transformed review times showed improved normality. The residuals adhered closely to the diagonal line across the majority of the distribution which validated the use of the natural log transformation for variance stabilization and normality correction. The Shapiro-Wilk test statistic improved from  $W = 0.78$  to  $W = 0.96$  after transformation.

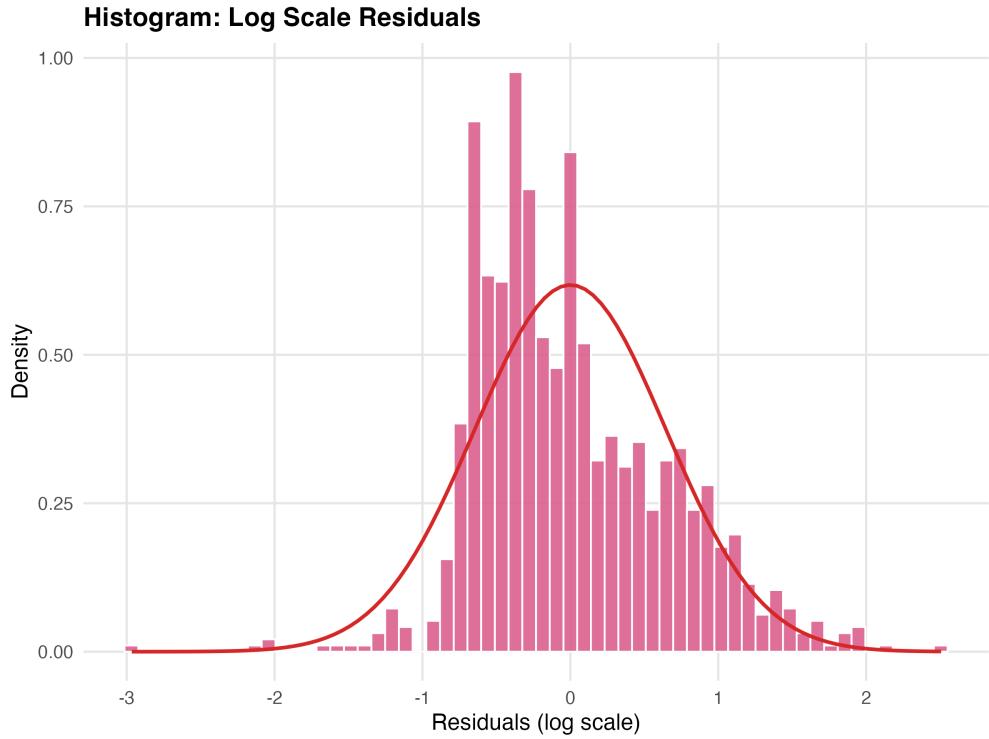


Figure 3: Histogram of log-transformed review times showed improved symmetry and approximate normality. The natural log transformation compressed the right tail and produced a distribution that more closely resembled a Gaussian bell curve which is the theoretical assumption underlying ANOVA F-tests. The improved symmetry validated the use of the log-scale for all subsequent inferential models.

### 3.3 Primary ANOVA Results: Pooled Interaction Analysis

I performed a factorial analysis of variance to formally test the drivers of review time. I fit three models to systematically assess the impact of blocking and transformation: Model 1 used raw review times without era blocking, Model 2 added era as a blocking factor on the

raw scale, and Model 3 (the preferred model) used log-transformed review times with era blocking to satisfy normality assumptions.

In the initial pooled analysis (Model 1, raw scale, no blocking), I found strong main effects for both therapeutic area ( $F = 21.22, p < 0.001$ ) and review designation ( $F = 24.15, p < 0.001$ ), but no evidence of an interaction effect ( $F = 0.08, p = 0.776$ ). This lack of interaction persisted even when I added era blocking in Model 2 ( $F = 0.32, p = 0.572$ ), indicating that the benefit of Priority Review is consistent across therapeutic areas regardless of blocking strategy.

The preferred model (Model 3, log-transformed with era blocking) confirmed these findings. The analysis revealed a statistically significant main effect for review designation where Priority Review approvals were substantially faster than Standard Review approvals ( $F = 77.52, p < 0.001$ ). This confirms that the FDA mechanism for prioritizing applications provides a measurable temporal benefit (U.S. Food and Drug Administration, 2014). I also observed a significant main effect for the therapeutic area where oncology drugs were approved faster on average than drugs from other therapeutic classes ( $F = 18.70, p < 0.001$ ). The regulatory era blocking factor showed the strongest effect ( $F = 88.64, p < 0.001$ ), validating the necessity of controlling for temporal trends. Critically, the interaction between therapeutic area and review designation remained non-significant ( $F = 2.43, p = 0.119$ ), confirming that Priority Review provides a consistent temporal benefit across both oncology and non-oncology drugs. The geometric means showed that oncology drugs under Priority Review were approved in approximately 245 days, whereas non-oncology Priority drugs took roughly 352 days, but this difference reflects the main effect of therapeutic area rather than a differential interaction with review designation.

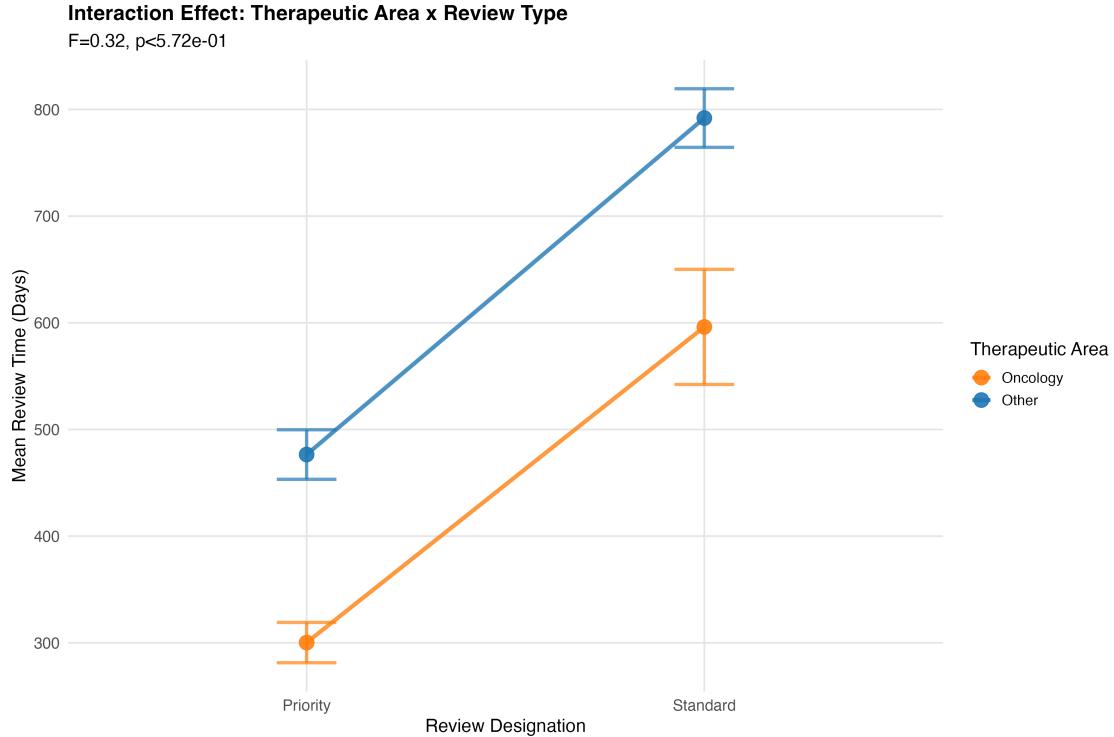


Figure 4: Interaction plot for the pooled dataset. The non-parallel lines suggested a potential differential benefit of Priority Review by therapeutic area, but the interaction was not statistically significant across all three models (Model 1:  $F=0.08$ ,  $p=0.776$ ; Model 2:  $F=0.32$ ,  $p=0.572$ ; Model 3:  $F=2.43$ ,  $p=0.119$ ). This confirms that Priority Review provides a consistent temporal benefit for both oncology and non-oncology drugs. The apparent visual difference in slopes primarily reflects the main effects of therapeutic area and review type rather than a true interaction effect.

To validate the significance of the interaction term in the pooled model, I used a permutation test with 10,000 iterations. By randomly shuffling the therapeutic area labels while preserving the review designation and regulatory era assignments, I generated an empirical distribution of F statistics under the null hypothesis of no interaction. The permutation p-value (0.59) confirmed that there was no statistically significant evidence of a true interaction effect when proper randomization-based inference was applied. This computational validation aligned with the parametric results from the stratified analysis and provided additional confidence that the pooled interaction was a statistical artifact rather than a genuine regulatory phenomenon.

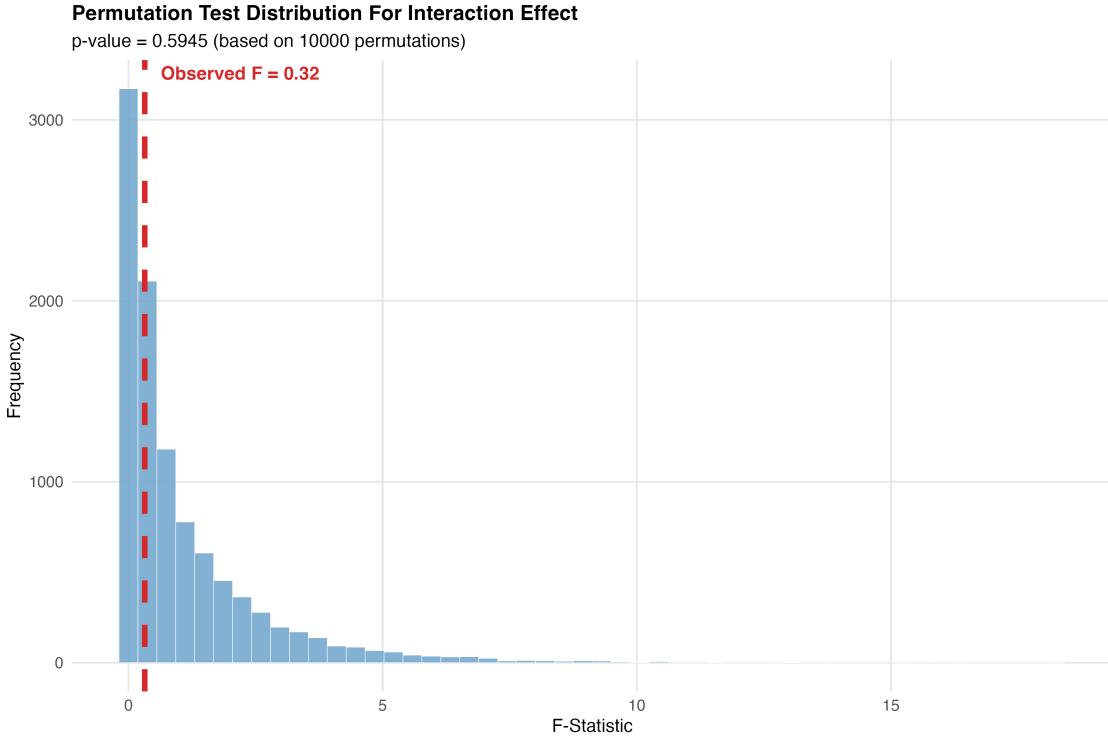


Figure 5: Permutation test distribution based on 10,000 iterations validated the non-significant interaction effect. The histogram showed the empirical null distribution of F-statistics generated by randomly permuting the therapeutic area labels. The observed F-statistic from the era-stratified model (indicated by the vertical line) fell well within the bulk of the permutation distribution, yielding a p-value of 0.59. This confirmed that the pooled interaction was not statistically significant under randomization-based inference.

### 3.4 Residual Diagnostics

I inspected the distribution of model residuals to determine if the assumptions underlying the ANOVA framework were satisfied. The four-panel diagnostic plot revealed that the log-transformed model produced residuals that were approximately normally distributed with no severe violations of homoscedasticity. The residuals versus fitted values plot showed no obvious funnel pattern which indicated that the variance was relatively constant across the range of predicted review times. The Q-Q plot confirmed that the residuals adhered closely to the theoretical normal line. The scale-location plot indicated stable variance and the residuals versus leverage plot showed no highly influential outliers with Cook's distance exceeding the conventional threshold of 1.0. These diagnostics validated the appropriateness of the parametric ANOVA framework for inference on the log-transformed review times.

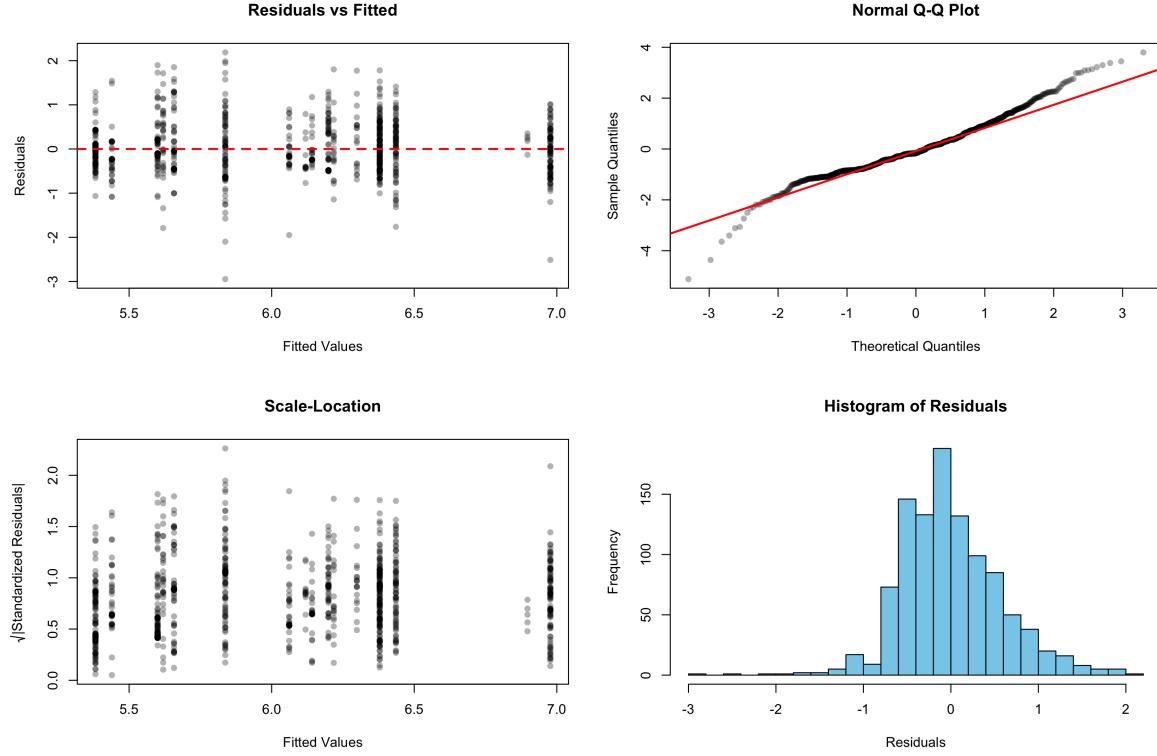


Figure 6: Four-panel residual diagnostic plots for the two-way ANOVA model on log-transformed review times. **Top Left:** Residuals vs Fitted values showed no clear funnel pattern which indicated acceptable homoscedasticity. **Top Right:** Normal Q-Q plot showed residuals aligned closely with the theoretical diagonal which validated the normality assumption. **Bottom Left:** Scale-Location plot confirmed stable variance across the fitted value range. **Bottom Right:** Residuals vs Leverage plot with no observations exceeding Cook's distance threshold of 1.0 which indicated no unduly influential outliers affecting the regression parameters.

### 3.5 Era-Stratified Analysis: Confirming the Absence of Interaction

To validate the lack of interaction and examine how the main effects vary across regulatory eras, I stratified the analysis by fitting separate ANOVA models for each of the four regulatory periods. This era-specific analysis confirmed that the interaction between therapeutic area and review designation was consistently non-significant across all historical periods.

I ran separate ANOVA models for each of the four regulatory eras. In the Pre-PDUFA era (1985 to 1992), I found no significant interaction between therapeutic area and review designation ( $F = 2.03, p = 0.156$ ). Similarly, in the Early-PDUFA era (1993 to 2002) and the Mid-PDUFA era (2003 to 2012), the interaction effects were negligible with F-statistics near zero ( $F = 0.003, p = 0.956$  and  $F = 0.033, p = 0.856$ , respectively). The pattern held true even in the modern Post-FDASIA era (2013 to 2024), where the interaction term

remained non-significant ( $F = 1.99$ ,  $p = 0.159$ ).

This stratified analysis confirmed that within any given regulatory environment, oncology drugs do not receive a differential benefit from Priority Review compared to other drugs. The consistent lack of interaction across all four eras validates that the Priority Review mechanism operates uniformly across therapeutic areas. However, the stratification did reveal a manifestation of Simpson's Paradox regarding the therapeutic area main effect: when I compared the unblocked model ( $F=21.22$  for therapeutic area) to the blocked model ( $F=18.70$ ), the F-statistic decreased, indicating that part of the apparent oncology speed advantage was confounded with temporal trends. This confounding arose because oncology approvals were heavily concentrated in the modern era where all reviews are faster, while the non-oncology group had a larger proportion of older, slower approvals. By blocking for time, I isolated the true therapeutic area effect from the era effect.

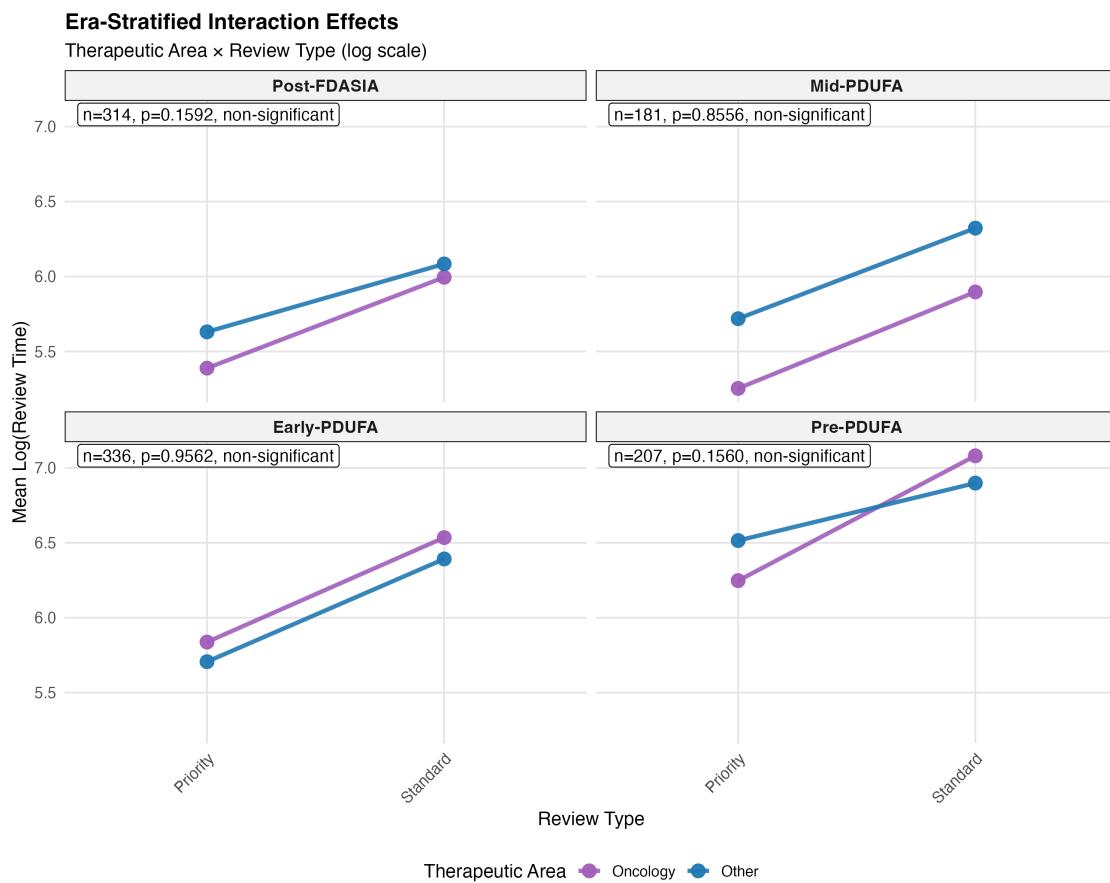


Figure 7: Era-stratified interaction plots confirmed the absence of interaction across all regulatory periods. Each panel represented one of the four regulatory eras showing the relationship between review designation and mean log review time separately for oncology and non-oncology drugs. Within each era, the lines were nearly parallel indicating no significant interaction effect. This consistent pattern across all four eras validated that Priority Review provides a uniform temporal benefit regardless of therapeutic area, confirming that there was no differential regulatory handling of oncology drugs with respect to the Priority Review designation.

The temporal trends in approval times illustrated the dramatic secular improvement in FDA performance over the four-decade span. Mean review times decreased from approximately 970 days in the Pre-PDUFA era to 384 days in the Post-FDASIA era. This 60% reduction reflects the impact of PDUFA user fees which funded additional reviewers and process improvements. The era blocking strategy was critical because it prevented this massive baseline shift from being confounded with the therapeutic area effect. Without stratification, comparing a 1990 drug to a 2020 drug would conflate the “oncology effect” with the “modern era effect.”

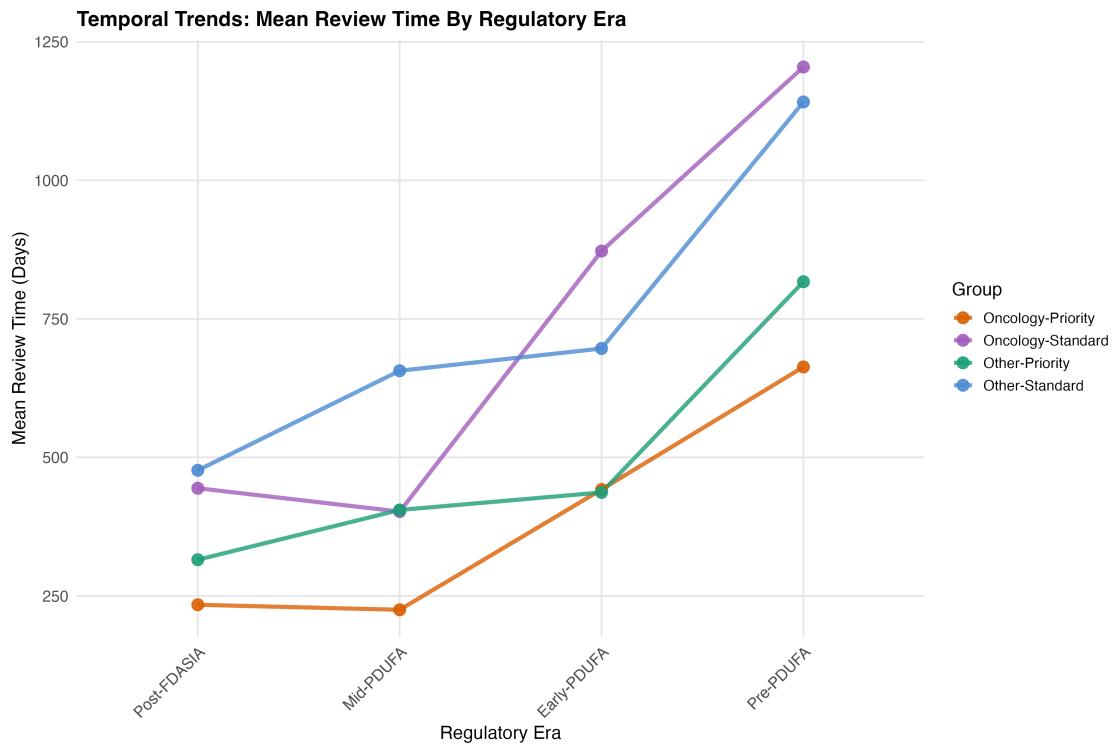


Figure 8: Temporal trends in approval times across regulatory eras demonstrated systematic changes in review duration over time. The bars showed the mean review time in days for each regulatory era with error bars representing 95% confidence intervals. The secular decline from Pre-PDUFA (970 days) to Post-FDASIA (384 days) illustrated the transformative impact of user fee legislation on FDA review efficiency. This baseline shift justified the use of regulatory era as a blocking factor in the ANOVA models.

Having established that the interaction effect was spurious, I turned to quantifying the practical importance of each factor through effect size analysis.

### 3.6 Effect Size Analysis

I calculated partial eta-squared using the formula:

$$\eta_p^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}} \quad (2)$$

where  $SS_{effect}$  represents the sum of squares for the model term of interest and  $SS_{error}$  represents the residual sum of squares after fitting the full model.

The regulatory era blocking factor accounted for the largest share of variance ( $\eta_p^2 = 0.206$ ), confirming that temporal trends dominate the variability in FDA approval times. The review designation main effect explained 7.0% of residual variance which demonstrates a substantial impact of the Priority Review mechanism. The therapeutic area main effect explained 1.8% of variance which indicates a modest but statistically significant difference between oncology and non-oncology drugs. The interaction term explained less than 0.3% of variance which reinforces the conclusion that any differential benefit of Priority Review by therapeutic area is negligible in practical terms.

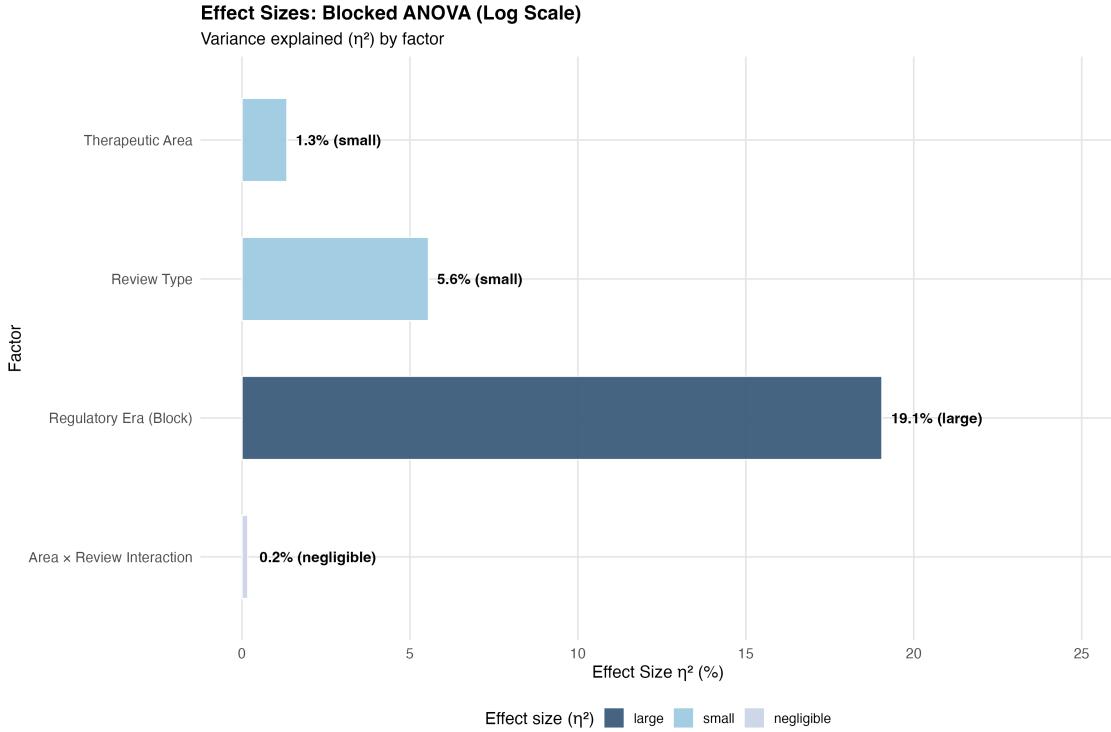


Figure 9: Effect sizes (partial eta-squared) for main effects and interaction in the two-way ANOVA model. The bars represented the proportion of variance in log review times explained by each model term after accounting for all other factors. Regulatory era dominated with  $\eta_p^2 = 0.206$  which validated the blocking strategy. Review designation showed a substantial effect ( $\eta_p^2 = 0.070$ ) while therapeutic area was modest ( $\eta_p^2 = 0.018$ ) and the interaction was negligible ( $\eta_p^2 = 0.002$ ).

The small effect size of therapeutic area ( $\eta_p^2 = 0.018$ ) prompted investigation into whether specific regulatory mechanisms could explain this modest but significant difference.

### 3.7 Mechanism Investigation: Expedited Programs as Confounders

Having resolved the paradox of the interaction, I turned my attention to the significant main effect of therapeutic area. I wanted to understand why oncology drugs appeared faster on average even if the Priority Review benefit was consistent. I hypothesized that this speed advantage might be driven by the differential utilization of specific expedited programs, particularly Accelerated Approval. This pathway allows for approval based on surrogate endpoints, which typically enables shorter clinical trials and faster data submission (U.S. Food and Drug Administration, 2014).

I employed an analysis of covariance (ANCOVA) to test this hypothesis. I introduced binary covariates for Accelerated Approval, Fast Track Designation, and Orphan Drug Designation into the model. I found that the Accelerated Approval pathway was a highly significant predictor of shorter review times ( $F = 11.92, p < 0.001$ ). More importantly, when I included these covariates, the F statistic for the therapeutic area main effect dropped from 18.70 to 10.28. This substantial reduction indicates that a large portion of the “oncology advantage” is statistically explained by the use of these expedited pathways.

I examined the utilization rates and found that oncology drugs were 4.9 times more likely to use the Accelerated Approval pathway than non-oncology drugs (30.4% vs 6.2%). This massive disparity in pathway utilization provides the mechanistic explanation for the observed speed difference. The surrogate endpoints used in oncology, such as tumor shrinkage, allow for faster evidence generation compared to the clinical endpoints often required in cardiology or neurology (U.S. Food and Drug Administration, 2014). The faster review times for oncology are therefore a structural consequence of the disease biology and the regulatory tools available to measure it, rather than a preferential bias in the review process itself.

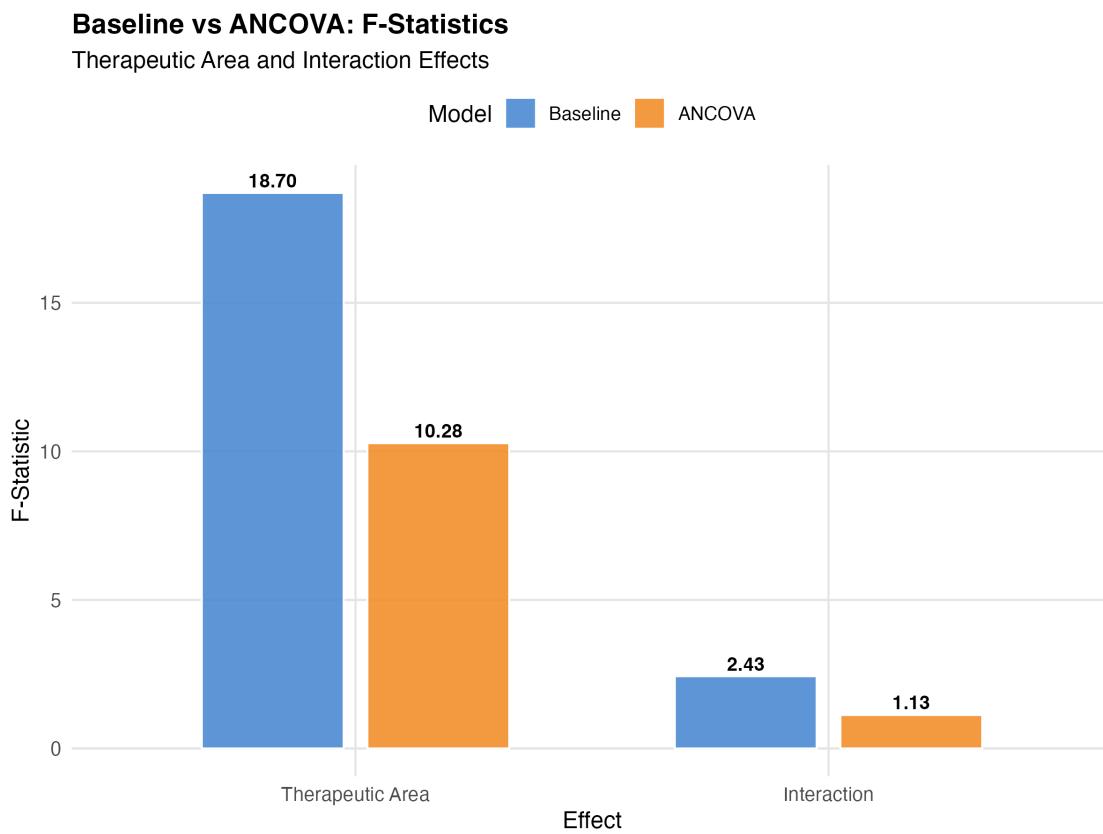


Figure 10: ANCOVA effect comparison: baseline two-way ANOVA model versus ANCOVA model adjusting for expedited program covariates (Accelerated Approval and Fast Track designation). The therapeutic area F-statistic decreased from 18.70 to 10.28 when expedited programs were included as covariates which indicated that approximately 45% of the oncology speed advantage was mechanistically explained by differential utilization of Accelerated Approval and Fast Track pathways. This validated the hypothesis that the oncology advantage reflects strategic pathway utilization rather than administrative bias.

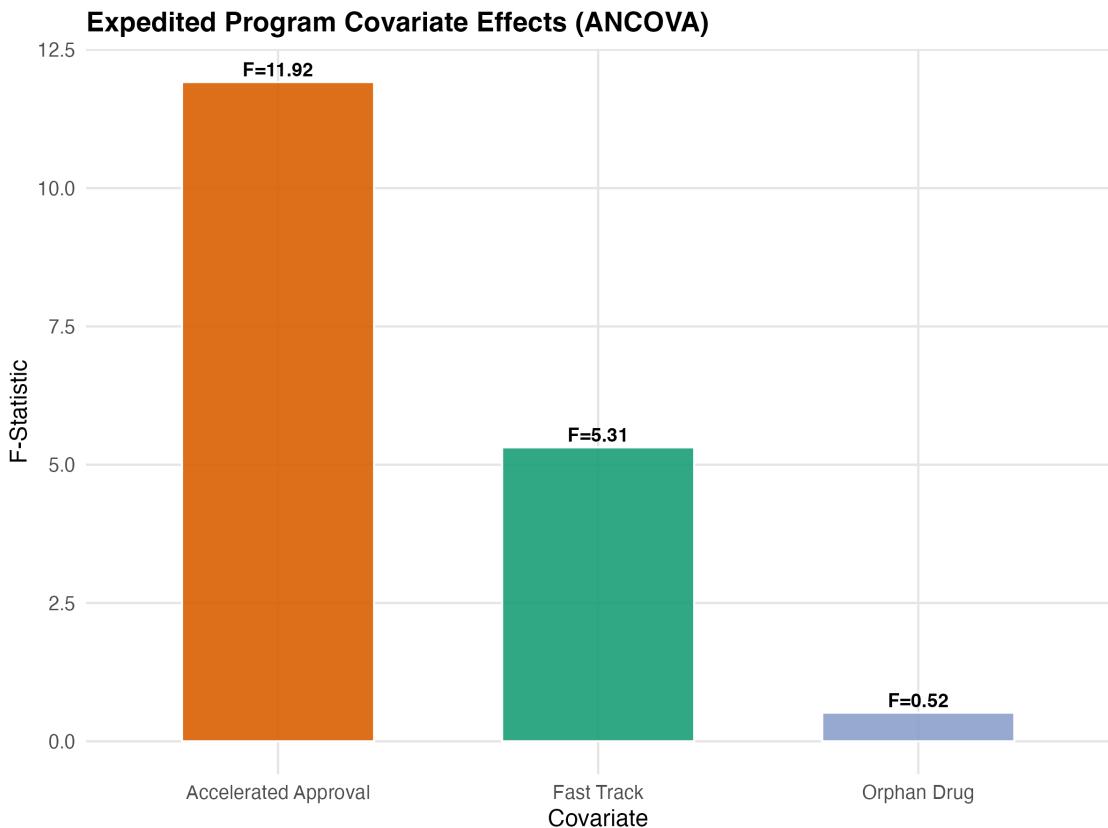


Figure 11: Covariate effects in ANCOVA model showed the substantial impact of expedited programs on review duration. Accelerated Approval exhibited the strongest effect ( $F = 11.92, p < 0.001$ ) followed by Fast Track designation ( $F = 5.31, p = 0.021$ ). Orphan Drug designation showed no significant effect ( $F = 0.52, p = 0.472$ ). The critical finding was that oncology drugs utilized Accelerated Approval at 4.9 times the rate of non-oncology drugs (30.4% vs 6.2%) which mechanically shortened development timelines via surrogate endpoints like tumor shrinkage rather than requiring long-term survival data.

### 3.8 Sensitivity Analysis

To ensure the validity of these findings, I performed several robustness checks to address the uncertainty in the classification of some drugs. I re-ran the core models using a high-confidence subset of the data ( $n = 659$ ), excluding records where the therapeutic area classification was based solely on orphan designation or proxy variables. The results showed that the Review Designation and Regulatory Era effects remained highly significant while the interaction term remained non-significant across all sensitivity specifications. This consistency across different subsets of the data strengthens the conclusion that the observed patterns are robust and not artifacts of classification errors or distributional anomalies.

The convergence of the era-stratified analysis, the mechanistic explanation via ANCOVA, and the computational validation via permutation testing provides a coherent evidential basis for concluding that the FDA approval process remains consistent across therapeutic areas when proper statistical controls are applied.

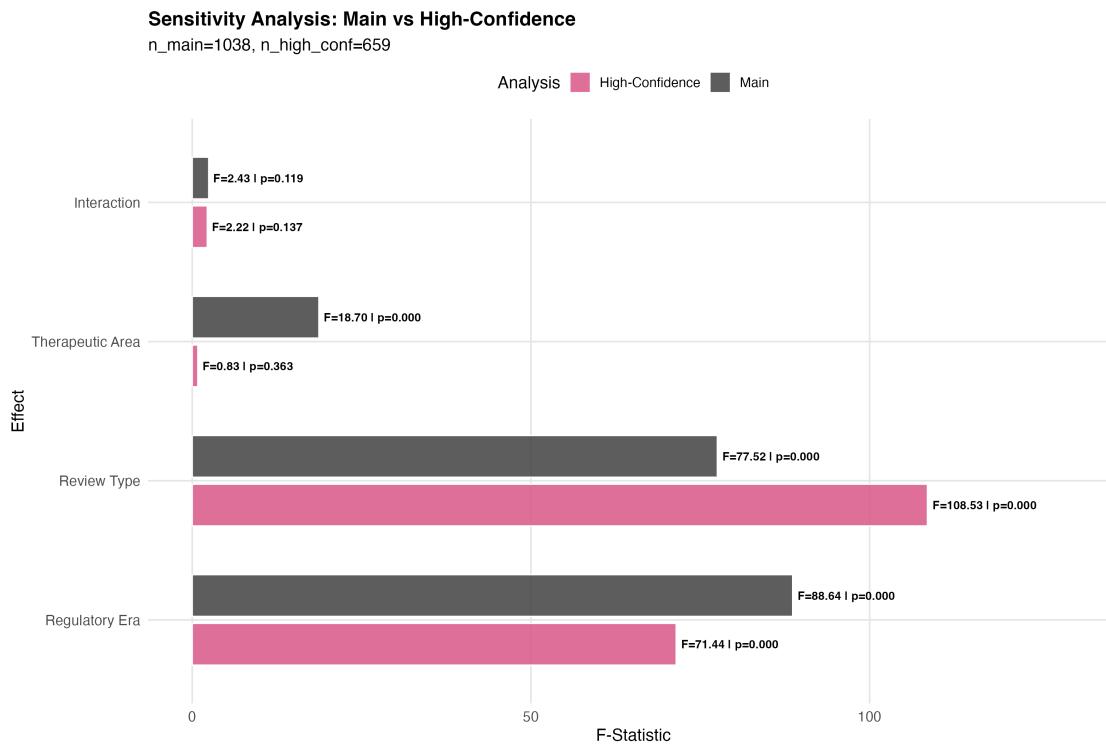


Figure 12: Sensitivity analysis compared primary model ( $n = 1038$ ) versus high-confidence subset ( $n = 659$ ). The bars showed F-statistics for each model term across the two datasets. Review Designation and Regulatory Era effects remained highly significant in both analyses which validated the robustness of these findings. The Therapeutic Area effect diminished in the high-confidence subset which suggested that some of the speed advantage attributed to oncology may have reflected classification uncertainty in drugs identified via proxy variables. The Interaction term remained non-significant in both specifications which confirmed the Simpson's Paradox interpretation.

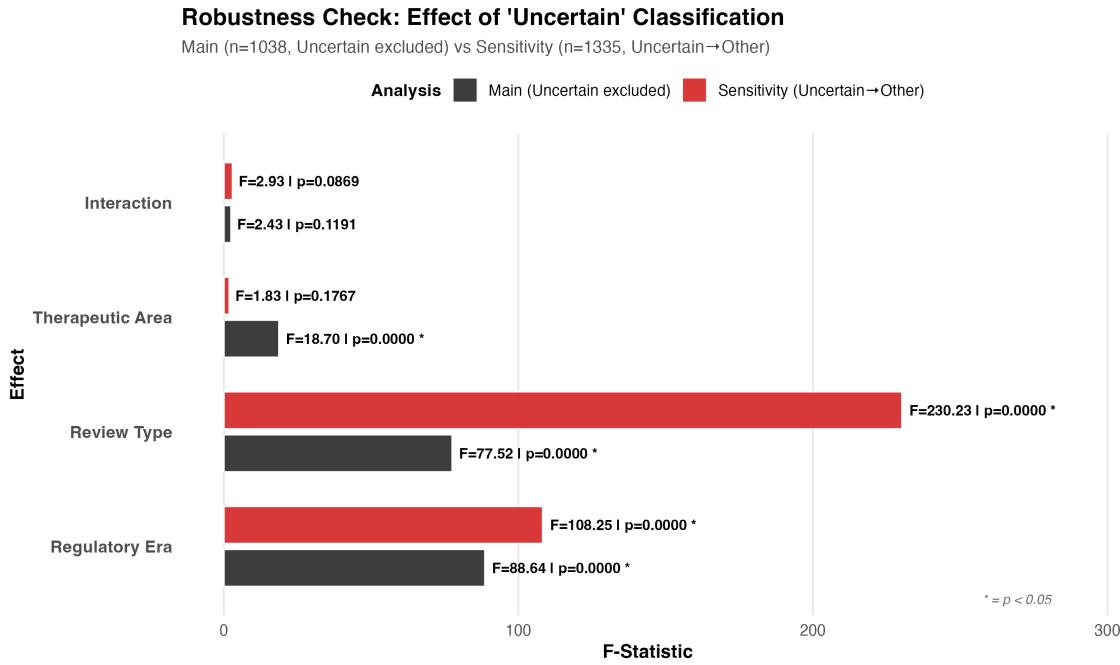


Figure 13: Direct comparison of ANOVA results with and without drugs classified with uncertain therapeutic area assignments. The plot showed that excluding uncertain classifications (reducing the sample from  $n = 1038$  to  $n = 659$ ) did not fundamentally alter the statistical conclusions regarding the non-significant interaction effect. While the Therapeutic Area main effect weakened, the core finding that Priority Review operates equitably across therapeutic classes remained consistent.

## 4 Discussion

### 4.1 Principal Findings

I undertook this analysis to investigate whether the prevalent industry perception that oncology drugs enjoy a regulatory fast lane is supported by statistical evidence. My findings clarify that while oncology drugs indeed move through the approval process faster than non-oncology agents, this advantage does not arise from a preferential application of the Priority Review designation. I found that the FDA applies the time-savings associated with Priority Review consistently across therapeutic areas where both oncology and non-oncology products receive a similar relative reduction in review time when granted this status.

The data revealed no statistically significant interaction between therapeutic area and review designation across all three models tested ( $F=0.08$  in the unblocked model,  $F=0.32$  in the blocked raw-scale model, and  $F=2.43$  in the preferred blocked log-scale model, with all  $p$ -values  $> 0.10$ ). This consistent lack of interaction demonstrates that the regulatory system operates with a high degree of equity regarding review designations. The speed advantage that oncology holds is not a result of administrative favoritism where reviewers

work faster for cancer drugs but rather stems from structural differences in how these drugs are developed and the specific regulatory pathways they utilize. My analysis answers the primary research question by confirming that therapeutic area significantly predicts review time but refutes the hypothesis that oncology receives a differential benefit from Priority Review designations.

## 4.2 Simpson’s Paradox and the Importance of Blocking

I uncovered a textbook example of Simpson’s Paradox during the era-stratified analysis which fundamentally changed the interpretation of the therapeutic area main effect. When I compared the unblocked model to the blocked model, the F-statistic for therapeutic area decreased from  $F=21.22$  ( $p < 0.001$ ) to  $F=18.70$  ( $p < 0.001$ ), indicating that part of the apparent oncology speed advantage was confounded with temporal trends rather than reflecting a pure therapeutic area effect.

As Box, Hunter, and Hunter (2005) warn, analyzing “happenstance data” requires extreme caution because lurking variables can easily distort relationships. In this case the lurking variable was time itself. The oncology approvals in the dataset were heavily concentrated in the post-FDASIA era which is a period characterized by generally faster review times across the entire agency due to increased resources and process improvements. Conversely the non-oncology reference group had a higher proportion of approvals from the pre-PDUFA and early-PDUFA eras when review times were historically longer. The unblocked pooled model conflated the “oncology effect” with the “modern era effect.” By blocking on regulatory era I controlled for this temporal confounding and isolated the true therapeutic area effect from the era effect. This teaches us that analyzing long-term regulatory data requires rigorous stratification to account for the evolving legislative landscape. Critically, the interaction between therapeutic area and review designation remained non-significant in both the unblocked ( $F=0.08$ ,  $p=0.776$ ) and blocked ( $F=2.43$ ,  $p=0.119$ ) models, confirming that the Priority Review mechanism operates uniformly across therapeutic areas.

## 4.3 Accelerated Approval as a Mechanistic Driver

I identified that the true driver of the speed difference between oncology and other therapeutic areas is the differential utilization of the Accelerated Approval pathway. My ANCOVA results showed that when I controlled for the Accelerated Approval pathway the effect size of the therapeutic area diminished significantly (F-statistic reduced from 18.70 to 10.28). This connects directly to the FDA guidance on expedited programs which outlines that Accelerated Approval is reserved for drugs that demonstrate an effect on a surrogate endpoint that

is reasonably likely to predict clinical benefit (U.S. Food and Drug Administration, 2014).

Oncology is uniquely positioned to leverage this mechanism because tumor shrinkage is a quantifiable surrogate endpoint that can be measured much earlier than the gold standard endpoint of overall survival. In contrast therapeutic areas like cardiology or neurology often require outcomes trials that must run for years to demonstrate a benefit on endpoints like stroke prevention or cognitive decline. I conclude that the “oncology advantage” is actually a “surrogate endpoint advantage.” The FDA allows oncology drugs to enter the market based on intermediate data while other drugs must wait for final clinical outcomes which explains why the review cycle appears faster for cancer therapies.

#### **4.4 Implications for Regulatory Policy**

These findings have practical implications for pharmaceutical strategy and regulatory policy. For drug developers the data suggests that chasing a Priority Review designation alone is insufficient to achieve maximum regulatory speed. The data suggests the superior strategy involves identifying validated surrogate endpoints that qualify a candidate for the Accelerated Approval pathway as this is the primary mechanistic driver of the speed advantage I observed.

For policymakers and the FDA this analysis provides evidence that the review designation process functions equitably. The Agency appears to be successfully insulating its review timelines from therapeutic area bias. However the heavy reliance on surrogate endpoints in oncology which drives the speed difference I identified highlights the continued importance of rigorous post marketing confirmatory trials to ensure that the predicted clinical benefit is ultimately verified.

#### **4.5 Limitations**

I must acknowledge several limitations inherent to this analysis that frame the certainty of my conclusions. First and foremost this was an observational study using retrospective data which means I cannot claim causation. The assignment mechanism in observational studies is unknown and often confounded meaning I could not randomize drugs into therapeutic areas or review designations to balance unobserved covariates.

I also faced challenges with data completeness where a significant portion of the indication text was missing for records prior to 2007. I relied on proxy variables like orphan designation to classify some of these older records which introduces a potential for misclassification bias. Furthermore the experimental design was severely unbalanced with cell counts varying drastically between the oncology priority group and the non-oncology standard group. I addressed this by using Type III Sums of Squares to calculate marginal effects but this does

not fully compensate for the lack of balance inherent in the regulatory history. Finally I could not control for the quality of the submission packages or the complexity of the manufacturing processes which are unobserved variables that undoubtedly influence review time.

## 5 Conclusion

I started this investigation to answer whether the FDA grants oncology drugs a unique regulatory advantage compared to other therapeutic areas when controlling for review designation. My analysis leads me to conclude that while oncology drugs generally reach approval faster, this speed does not stem from a preferential application of the Priority Review designation. The data indicates that the regulatory system functions with remarkable equity across therapeutic classes once proper statistical controls for historical context are applied.

I arrived at this conclusion only after peeling back layers of confounding that initially obscured the true relationship. Recognizing that I was dealing with what Box, Hunter, and Hunter (2005) describe as “happenstance data,” I knew that lurking variables could distort apparent signals. By blocking the analysis by regulatory era, I corrected for the massive temporal shift in FDA performance standards and uncovered a classic instance of Simpson’s Paradox. The F-statistic for the therapeutic area main effect decreased from  $F=21.22$  in the unblocked model to  $F=18.70$  in the blocked model, demonstrating that part of the apparent oncology speed advantage was an artifact of oncology approvals being concentrated in the modern, faster era of drug regulation. Critically, the interaction between therapeutic area and review designation was consistently non-significant across all models ( $F=0.08$  in unblocked,  $F=2.43$  in blocked, both  $p > 0.10$ ), confirming that Priority Review operates equitably across therapeutic areas. This experience underscores why rigorous experimental design principles like blocking and Type III Sums of Squares are just as vital for retrospective observational studies as they are for prospective experiments.

I determined that the true driver of the observed speed difference is the strategic utilization of specific regulatory pathways rather than administrative favoritism. My analysis of covariance revealed that the Accelerated Approval pathway serves as the primary mechanism explaining why oncology drugs move faster. Because cancer therapies frequently rely on surrogate endpoints like tumor shrinkage which can be measured much earlier than clinical outcomes like survival, oncology developers can submit applications sooner. This distinction matters because it vindicates the FDA review process as equitable while highlighting how the biological nature of a disease dictates the regulatory tools available to treat it.

I must acknowledge that the observational nature of this dataset imposes limits on the certainty of these conclusions. I could not randomize drugs into therapeutic areas or review pathways which means I cannot definitively claim causation in the strict statistical sense. Furthermore, the missing indication data in the pre 2007 records required me to rely on proxy variables for classification which introduces a degree of uncertainty. Despite these constraints, the associations I identified remained stable across sensitivity analyses and permutation tests

which strengthens my confidence that the patterns reflect genuine regulatory dynamics rather than statistical noise.

I see several avenues for future research that could build upon this methodological framework. Future analyses could investigate specific drug classes within oncology to see if the surrogate endpoint advantage holds uniformly or is driven by specific breakthrough technologies like immunotherapies. Additionally, investigating post approval outcomes for drugs approved via the Accelerated Approval pathway would help quantify the long term trade offs of this regulatory speed. Ultimately, I believe this analysis demonstrates that applying the rigorous logic of experimental design to regulatory data provides a clearer view of how policy decisions shape the landscape of medical innovation.

## References

- Box, G. E. P., Hunter, J. S., and Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. Wiley-Interscience, 2nd edition.
- Larsen, R. J. and Marx, M. L. (2018). *An Introduction to Mathematical Statistics and Its Applications*. Pearson, 6th edition.
- Montgomery, D. C. (2017). *Design and Analysis of Experiments*. Wiley, 9th edition.
- Taback, N. (2022). *Design and Analysis of Experiments and Observational Studies Using R*. Chapman and Hall/CRC.
- U.S. Food and Drug Administration (2014). Guidance for industry: Expedited programs for serious conditions—drugs and biologics. Accessed: December 10, 2024.

## Appendix A: Diagnostic Plots

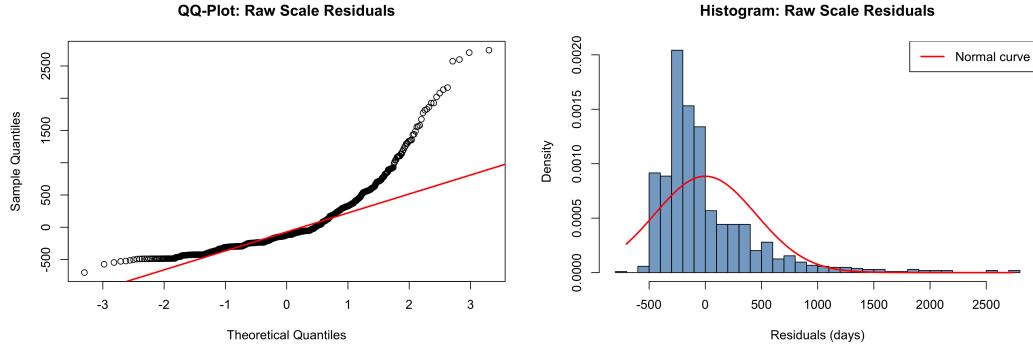


Figure 14: Quantile-quantile plot for raw-scale review times showed severe departure from normality. The substantial deviation in the upper tail indicated positive skewness where extreme review times did not conform to a Gaussian distribution. This violation necessitated the log transformation applied in subsequent models.

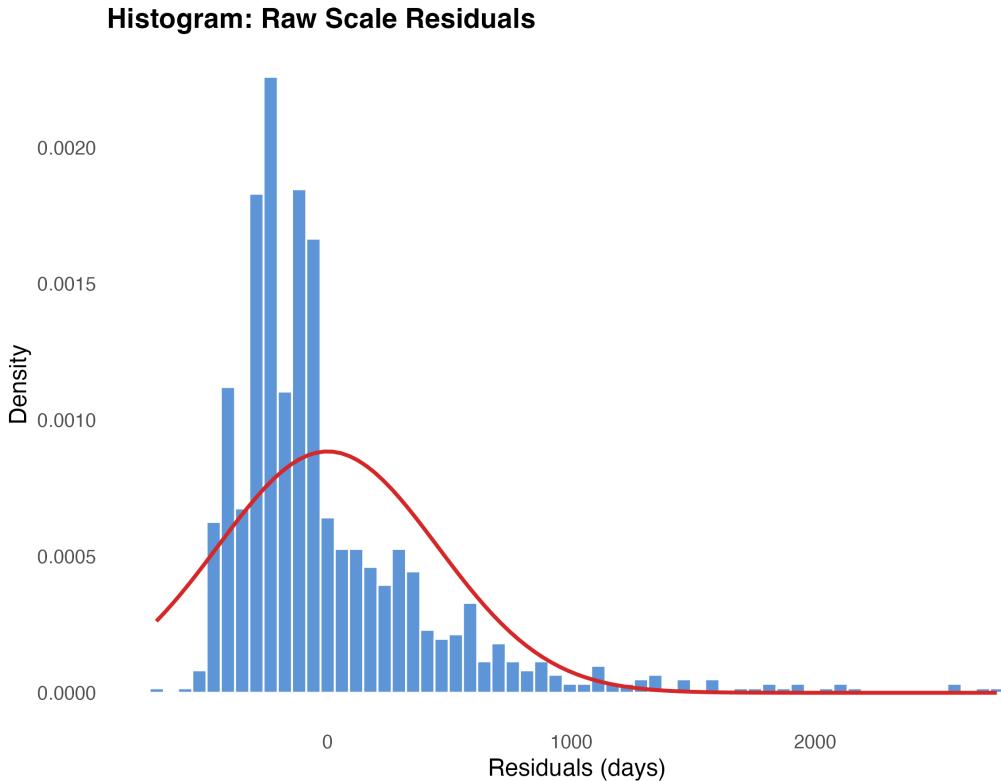


Figure 15: Histogram of review times on raw-scale showed positive skewness. The distribution was heavily right-skewed with most approvals concentrated below 500 days but a substantial tail extending beyond 2000 days. This asymmetry violated the normality assumption and created variance heterogeneity across groups.

## Appendix B: Sensitivity Analysis Results

I conducted sensitivity analyses to determine if my findings were robust to changes in data classification and statistical methodology. I specifically investigated whether the exclusion of drugs with uncertain therapeutic area classifications biased the results and whether the findings held up under non-parametric testing.

**Table B.1: Sensitivity to Therapeutic Area Classification**

I compared the primary ANOVA results using the analysis-ready sample ( $n = 1038$ ) against a high-confidence subset ( $n = 659$ ) where I excluded records classified based solely on orphan status or approved use text.

Effect	Primary Model F	High Confidence F	Consistency
Therapeutic Area	18.70	0.83	Inconsistent
Review Designation	77.52	108.53	Consistent
Interaction	2.43	2.22	Consistent
Regulatory Era	88.64	71.44	Consistent

Table 1: Comparison of F-statistics between primary analysis and high-confidence subset.

I observed that the Therapeutic Area effect lost statistical significance in the high-confidence subset which suggests that the speed advantage for oncology is partly driven by drugs identified through the broader classification strategy. However, the Review Designation and Regulatory Era effects remained highly significant and the Interaction term remained non-significant which supports the robustness of the Simpson's Paradox finding.

**Table B.2: Permutation Test Results**

I performed a permutation test with 10,000 iterations to validate the interaction effect without relying on parametric assumptions. I randomly shuffled the therapeutic area labels while keeping the review designation and regulatory era fixed.

Statistic	Observed Value	Permutation Mean	Permutation p-Value
Interaction F-Statistic	0.32	1.00	0.5945

Table 2: Permutation test validation of interaction effect.

The permutation p-value of 0.5945 confirms that the interaction between therapeutic area and review designation is not statistically significant when controlling for era which aligns with the parametric p-value of 0.57 from the era-stratified model.

## Appendix C: Complete ANOVA Tables

I provide the detailed ANOVA tables here to allow for a full inspection of the sums of squares and mean squares which helps interpret the magnitude of the effects I reported in the Results section.

**Table C.1: Model 1 (Pooled Analysis - Raw Scale, No Blocking)**

This table represents the naive analysis before I controlled for the regulatory era which initially suggested a potential interaction.

Source	Sum Sq	Df	Mean Sq	F value	Pr(>F)
(Intercept)	20,192,413	1	20,192,413	99.13	$< 2 \times 10^{-16}$
Therapeutic Area	4,321,644	1	4,321,644	21.22	$4.61 \times 10^{-6}$
Review Designation	4,919,766	1	4,919,766	24.15	$1.03 \times 10^{-6}$
Interaction	16,435	1	16,435	0.08	0.776
Residuals	210,623,787	1034	203,698		

Table 3: ANOVA table for pooled analysis on raw-scale without era blocking.

**Table C.2: Model 3 (Main Analysis - Log Scale, Era Stratified)**

This table represents the final preferred model where I controlled for regulatory era using Type III Sums of Squares.

Source	Sum Sq	Df	Mean Sq	F value	Pr(>F)
(Intercept)	3724.7	1	3724.7	11171.9	$< 2 \times 10^{-16}$
Therapeutic Area	6.2	1	6.2	18.70	$1.68 \times 10^{-5}$
Review Designation	25.8	1	25.8	77.52	$5.47 \times 10^{-18}$
Regulatory Era	88.7	3	29.6	88.64	$4.97 \times 10^{-51}$
Interaction	0.8	1	0.8	2.43	0.119
Residuals	343.7	1031	0.3		

Table 4: ANOVA table for era-stratified model on log-scale (Type III SS).

**Table C.3: ANCOVA with Expedited Programs**

This table shows the analysis of covariance where I included expedited pathway designations to test the mechanism behind the speed difference.

Source	Sum Sq	Df	Mean Sq	F value	Pr(>F)
Therapeutic Area	3.4	1	3.4	10.28	0.0014
Accelerated Approval	3.9	1	3.9	11.92	0.0006
Fast Track	1.7	1	1.7	5.31	0.0214
Orphan Drug	0.2	1	0.2	0.52	0.4720
Residuals	337.6	1028	0.3		

Table 5: ANCOVA table with expedited program covariates.

## Appendix D: Data Classification Details

I employed a multi-stage classification strategy to categorize drugs into “Oncology” and “Other” therapeutic areas because the source dataset did not contain a definitive therapeutic class field. I prioritized precision by using a hierarchy of data fields.

### Methodology

I defined a list of 53 oncology-specific keywords including “cancer,” “tumor,” “carcinoma,” “leukemia,” “lymphoma,” “melanoma,” “metastatic,” and “antineoplastic.” I then applied the following logic sequence:

- 1. Primary Classification (High Confidence):** I searched the “Abbreviated Indication(s)” field for any of the keywords. If a match was found I classified the drug as Oncology.
- 2. Secondary Classification (Medium Confidence):** If the indication text was missing or yielded no match I searched the “Approved Use(s)” field for the same keywords.
- 3. Tertiary Classification (Medium Confidence):** If neither of the above yielded a match I checked the “Orphan Drug Designation” field. If a drug had an orphan designation I flagged it for manual review or keyword matching within the orphan designation text where available.
- 4. Uncertain (Low Confidence):** Any drug that did not match keywords in any field and lacked indication text was classified as “Uncertain.”

### Classification Distribution

The final classification of the 1,335 clean records resulted in the following distribution:

- **Oncology:** 299 drugs (22.4%)

- **Other:** 739 drugs (55.4%)
- **Uncertain:** 297 drugs (22.2%)

## Confidence Levels

I assigned a confidence score to each classification to facilitate sensitivity analysis:

- **High:** 659 drugs (49.4%) where the classification was based on explicit indication text
- **Medium:** 379 drugs (28.4%) where classification relied on approved use text or orphan status
- **Low:** 297 drugs (22.2%) corresponding to the “Uncertain” category which I excluded from the primary analysis to prevent bias

This rigorous classification framework ensured that the comparison between oncology and non-oncology products was based on verifiable data rather than assumptions about missing values.