phase4b_expedited_programs.R

abdulbasir

2025-12-09

```r
# phase 4b: expedited programs analysis
#
# analyzing usage of FDA expedited programs (Accelerated Approval, Fast Track,
# Orphan Drug) to explain differential review times between therapeutic areas

# sourcing configuration and utilities
source("config.R")
source("utils.R")

# loading required libraries
suppressPackageStartupMessages({
  library(dplyr)
  library(readr)
})

print_section_header("Phase 4b: Expedited Programs Analysis")
```

```
##
## ========================================================================
## Phase 4b: Expedited Programs Analysis
## ========================================================================
```

```r
# 1. loading classified data
input_file = file.path(RESULTS_DIR, "fda_analysis_clean_classified.csv")
classified_data = load_csv(input_file)

# filtering to analysis-ready sample (exclude Uncertain)
analysis_data = classified_data %>%
  filter(therapeutic_area != "Uncertain")

cat(paste("Analysis-ready sample (excluding Uncertain): n =", nrow(analysis_data), "\n"))
```

```
## Analysis-ready sample (excluding Uncertain): n = 1038
```

```r
# 2. analyzing Accelerated Approval usage
acc_approval_table = table(
  analysis_data$therapeutic_area,
  analysis_data$`Accelerated Approval`
)

cat("\nAccelerated Approval by Therapeutic Area:\n")
```

```
##
## Accelerated Approval by Therapeutic Area:
```

```
print(addmargins(acc_approval_table))
```

```
##
##            N/A  No  Yes Yes (indication [A] only) Yes (indication [B] only)
##   Oncology  21 182   91                         2                         2
##   Other    161 530   46                         1                         1
##   Sum      182 712  137                         3                         3
##
##           Yes (indications [B] and [C] only)  Sum
##   Oncology                                 1  299
##   Other                                    0  739
##   Sum                                      1 1038
```

```
acc_approval_pct = prop.table(acc_approval_table, margin = 1) * 100
cat("Percentages (row-wise):\n")
```

```
## Percentages (row-wise):
```

```
print(round(acc_approval_pct, 2))
```

```
##
##             N/A    No   Yes Yes (indication [A] only) Yes (indication [B] only)
##   Oncology 7.02 60.87 30.43                      0.67                      0.67
##   Other   21.79 71.72  6.22                      0.14                      0.14
##
##           Yes (indications [B] and [C] only)
##   Oncology                               0.33
##   Other                                  0.00
```

```
# extracting key statistics
acc_oncology_yes = acc_approval_table["Oncology", "Yes"]
acc_oncology_total = sum(acc_approval_table["Oncology", ])
acc_oncology_pct_val = (acc_oncology_yes / acc_oncology_total) * 100

acc_other_yes = acc_approval_table["Other", "Yes"]
acc_other_total = sum(acc_approval_table["Other", ])
acc_other_pct_val = (acc_other_yes / acc_other_total) * 100

cat(sprintf("\nOncology: %d/%d (%.1f%%) received Accelerated Approval\n",
            acc_oncology_yes, acc_oncology_total, acc_oncology_pct_val))
```

```
##
## Oncology: 91/299 (30.4%) received Accelerated Approval
```

```
cat(sprintf("Other: %d/%d (%.1f%%) received Accelerated Approval\n",
            acc_other_yes, acc_other_total, acc_other_pct_val))
```

```
## Other: 46/739 (6.2%) received Accelerated Approval
```

```r
cat(sprintf("Ratio: Oncology is %.1fx more likely to receive Accelerated Approval\n",
            acc_oncology_pct_val / acc_other_pct_val))
```

```
## Ratio: Oncology is 4.9x more likely to receive Accelerated Approval
```

```r
# 3. analyzing Fast Track usage
fast_track_table = table(
  analysis_data$therapeutic_area,
  analysis_data$`Fast Track Designation`
)

cat("\nFast Track by Therapeutic Area:\n")
```

```
##
## Fast Track by Therapeutic Area:
```

```r
print(addmargins(fast_track_table))
```

```
##
##              N/A   No  Yes Yes (indication [A] only) Yes (indication [B] only)  Sum
##   Oncology    51  140  104                        3                         1  299
##   Other      322  297  119                        0                         1  739
##   Sum        373  437  223                        3                         2 1038
```

```r
fast_track_pct = prop.table(fast_track_table, margin = 1) * 100
cat("Percentages (row-wise):\n")
```

```
## Percentages (row-wise):
```

```r
print(round(fast_track_pct, 2))
```

```
##
##               N/A     No    Yes Yes (indication [A] only) Yes (indication [B] only)
##   Oncology 17.06  46.82  34.78                      1.00                      0.33
##   Other    43.57  40.19  16.10                      0.00                      0.14
```

```r
# extracting key statistics
ft_oncology_yes = fast_track_table["Oncology", "Yes"]
ft_oncology_total = sum(fast_track_table["Oncology", ])
ft_oncology_pct_val = (ft_oncology_yes / ft_oncology_total) * 100

ft_other_yes = fast_track_table["Other", "Yes"]
ft_other_total = sum(fast_track_table["Other", ])
ft_other_pct_val = (ft_other_yes / ft_other_total) * 100

cat(sprintf("\nOncology: %d/%d (%.1f%%) received Fast Track\n",
            ft_oncology_yes, ft_oncology_total, ft_oncology_pct_val))
```

```
##
## Oncology: 104/299 (34.8%) received Fast Track
```

```
cat(sprintf("Other: %d/%d (%.1f%%) received Fast Track\n",
            ft_other_yes, ft_other_total, ft_other_pct_val))
```

Other: 119/739 (16.1%) received Fast Track

```
cat(sprintf("Ratio: Oncology is %.1fx more likely to receive Fast Track\n",
            ft_oncology_pct_val / ft_other_pct_val))
```

Ratio: Oncology is 2.2x more likely to receive Fast Track

```
# 4. analyzing Orphan Drug Designation usage
analysis_data = analysis_data %>%
  mutate(
    orphan_binary = if_else(
      `Orphan Drug Designation` %in% c("Yes", "yes"),
      "Yes",
      "No"
    )
  )

orphan_table = table(
  analysis_data$therapeutic_area,
  analysis_data$orphan_binary
)

cat("\nOrphan Drug by Therapeutic Area:\n")
```

```
##
## Orphan Drug by Therapeutic Area:
```

```
print(addmargins(orphan_table))
```

```
##
##             No   Yes   Sum
##   Oncology  129  170   299
##   Other     471  268   739
##   Sum       600  438  1038
```

```
orphan_pct = prop.table(orphan_table, margin = 1) * 100
cat("Percentages (row-wise):\n")
```

Percentages (row-wise):

```
print(round(orphan_pct, 2))
```

```
##
##              No    Yes
##   Oncology 43.14 56.86
##   Other    63.73 36.27
```

```r
# extracting key statistics
orp_oncology_yes = orphan_table["Oncology", "Yes"]
orp_oncology_total = sum(orphan_table["Oncology", ])
orp_oncology_pct_val = (orp_oncology_yes / orp_oncology_total) * 100

orp_other_yes = orphan_table["Other", "Yes"]
orp_other_total = sum(orphan_table["Other", ])
orp_other_pct_val = (orp_other_yes / orp_other_total) * 100

cat(sprintf("\nOncology: %d/%d (%.1f%%) received Orphan Designation\n",
            orp_oncology_yes, orp_oncology_total, orp_oncology_pct_val))
```

```
##
## Oncology: 170/299 (56.9%) received Orphan Designation
```

```r
cat(sprintf("Other: %d/%d (%.1f%%) received Orphan Designation\n",
            orp_other_yes, orp_other_total, orp_other_pct_val))
```

```
## Other: 268/739 (36.3%) received Orphan Designation
```

```r
cat(sprintf("Ratio: Oncology is %.1fx more likely to receive Orphan Designation\n",
            orp_oncology_pct_val / orp_other_pct_val))
```

```
## Ratio: Oncology is 1.6x more likely to receive Orphan Designation
```

```r
# 5. creating summary table
summary_table = data.frame(
  program = c("Accelerated Approval", "Fast Track", "Orphan Drug"),
  oncology_pct = c(acc_oncology_pct_val, ft_oncology_pct_val, orp_oncology_pct_val),
  other_pct = c(acc_other_pct_val, ft_other_pct_val, orp_other_pct_val)
) %>%
  mutate(
    ratio = oncology_pct / other_pct,
    diff = oncology_pct - other_pct
  )

cat("\nSummary of expedited program usage:\n")
```

```
##
## Summary of expedited program usage:
```

```r
print(summary_table)
```

```
##                program oncology_pct other_pct    ratio     diff
## 1 Accelerated Approval     30.43478  6.224628 4.889414 24.21015
## 2           Fast Track     34.78261 16.102842 2.160029 18.67977
## 3          Orphan Drug     56.85619 36.265223 1.567788 20.59096
```

```r
max_ratio_program = summary_table$program[which.max(summary_table$ratio)]
max_ratio_value = max(summary_table$ratio)

cat(sprintf("\nLargest disparity: %s (%.1fx more likely for Oncology)\n",
            max_ratio_program, max_ratio_value))
```

```
##
## Largest disparity: Accelerated Approval (4.9x more likely for Oncology)
```

```r
# 6. saving results
output_file = file.path(RESULTS_DIR, "expedited_programs_summary.csv")
save_csv(summary_table, output_file)
```

```
## saving results to: /Users/abdulbasir/Downloads/Experimental AI/fda-oncology-approval-analysis/results
```

```r
cat("\nPhase 4b complete\n")
```

```
##
## Phase 4b complete
```