Patient-Reported Experience and Sentiment Analysis of Birth Control Medications: A Data-Driven Approach Using R

Banan Bashar Al-jarrah

2025-07-9

Introduction

The widespread use of online drug reviews provides a unique opportunity to explore patient experiences and concerns beyond clinical trials. In this project, I focused on reviews related to birth control medications to identify common negative sentiments, temporal complaints, and evaluate drug-specific ratings

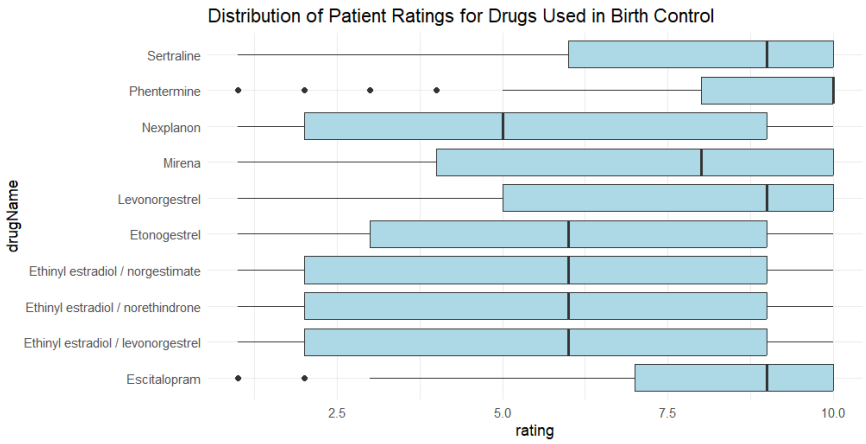
Data Import and Cleaning

data cleaning involved handling missing values (e.g., filling missing conditions), removing duplicates, and converting the date column into a standard date format. Drug and condition names were standardized for consistent analysis.

```
## # A tibble: 5 x 5
##   uniqueID drugName        condition                    review            rating
##      <dbl> <chr>           <chr>                        <chr>              <dbl>
## 1   163740 Mirtazapine     Depression                   "\"I&#039;ve tri~     10
## 2   206473 Mesalamine      Crohn's Disease, Maintenance "\"My son has Cr~      8
## 3   159672 Bactrim         Urinary Tract Infection      "\"Quick reducti~      9
## 4    39293 Contrave        Weight Loss                  "\"Contrave comb~      9
## 5    97768 Cyclafem 1 / 35 Birth Control                "\"I have been o~      9
```

Exploring Conditions and Drugs

Among the most frequently reviewed drugs for birth control were Lo Loestrin Fe, Ortho Tri-Cyclen, and Nexplanon. The analysis also highlighted the dominance of birth control as a reviewed condition in the dataset.



ANOVA Test for Rating Differences Among Drugs

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## drugName       4   1026   256.6   23.76 <2e-16 ***
## Residuals   3825  41310    10.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

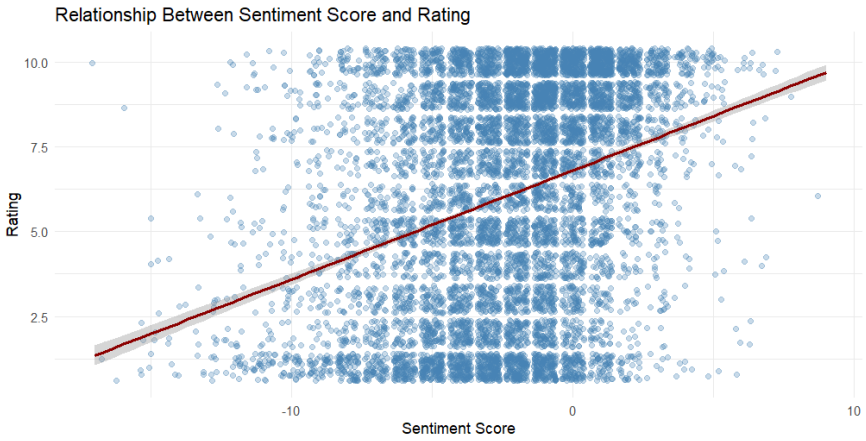
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = rating ~ drugName, data = review_top5)
##
## $drugName
##                                                                    diff
## Ethinyl estradiol / norgestimate-Ethinyl estradiol / norethindrone  0.09191261
## Etonogestrel-Ethinyl estradiol / norethindrone                      0.28330346
## Levonorgestrel-Ethinyl estradiol / norethindrone                    1.36594389
## Nexplanon-Ethinyl estradiol / norethindrone                        -0.04890715
## Etonogestrel-Ethinyl estradiol / norgestimate                       0.19139086
## Levonorgestrel-Ethinyl estradiol / norgestimate                     1.27403128
## Nexplanon-Ethinyl estradiol / norgestimate                         -0.14081976
## Levonorgestrel-Etonogestrel                                         1.08264042
## Nexplanon-Etonogestrel                                             -0.33221062
## Nexplanon-Levonorgestrel                                           -1.41485104
##                                                                        lwr
## Ethinyl estradiol / norgestimate-Ethinyl estradiol / norethindrone -0.4216037
## Etonogestrel-Ethinyl estradiol / norethindrone                     -0.1440060
## Levonorgestrel-Ethinyl estradiol / norethindrone                    0.9026364
## Nexplanon-Ethinyl estradiol / norethindrone                        -0.5154926
## Etonogestrel-Ethinyl estradiol / norgestimate                      -0.2882584
## Levonorgestrel-Ethinyl estradiol / norgestimate                     0.7620509
## Nexplanon-Ethinyl estradiol / norgestimate                         -0.6557683
## Levonorgestrel-Etonogestrel                                         0.6571781
## Nexplanon-Etonogestrel                                             -0.7612402
## Nexplanon-Levonorgestrel                                           -1.8797454
##                                                                        upr
```

```
## Ethinyl estradiol / norgestimate-Ethinyl estradiol / norethindrone  0.60542896
## Etonogestrel-Ethinyl estradiol / norethindrone                      0.71061296
## Levonorgestrel-Ethinyl estradiol / norethindrone                    1.82925140
## Nexplanon-Ethinyl estradiol / norethindrone                         0.41767830
## Etonogestrel-Ethinyl estradiol / norgestimate                       0.67104012
## Levonorgestrel-Ethinyl estradiol / norgestimate                     1.78601162
## Nexplanon-Ethinyl estradiol / norgestimate                          0.37412878
## Levonorgestrel-Etonogestrel                                         1.50810279
## Nexplanon-Etonogestrel                                              0.09681893
## Nexplanon-Levonorgestrel                                           -0.94995664
##                                                                        p adj
## Ethinyl estradiol / norgestimate-Ethinyl estradiol / norethindrone 0.9884448
## Etonogestrel-Ethinyl estradiol / norethindrone                     0.3681842
## Levonorgestrel-Ethinyl estradiol / norethindrone                   0.0000000
## Nexplanon-Ethinyl estradiol / norethindrone                        0.9985389
## Etonogestrel-Ethinyl estradiol / norgestimate                      0.8123365
## Levonorgestrel-Ethinyl estradiol / norgestimate                    0.0000000
## Nexplanon-Ethinyl estradiol / norgestimate                         0.9455003
## Levonorgestrel-Etonogestrel                                        0.0000000
## Nexplanon-Etonogestrel                                             0.2145215
## Nexplanon-Levonorgestrel                                           0.0000000
```

ANOVA results revealed a statistically significant difference in patient ratings across different birth control drugs ($p < 0.001$), indicating that user satisfaction varies by medication.

Sentiment Analysis Preparation

By tokenizing the reviews and using the Bing lexicon, I assigned a sentiment score to each review. A strong positive correlation was found between sentiment score and user rating, validating the sentiment approach



The scatter plot shows a clear positive trend between sentiment score and user rating. Reviews with more positive words tend to have higher ratings. This supports the idea that patient language sentiment is aligned with their numerical feedback.

Correlation Test

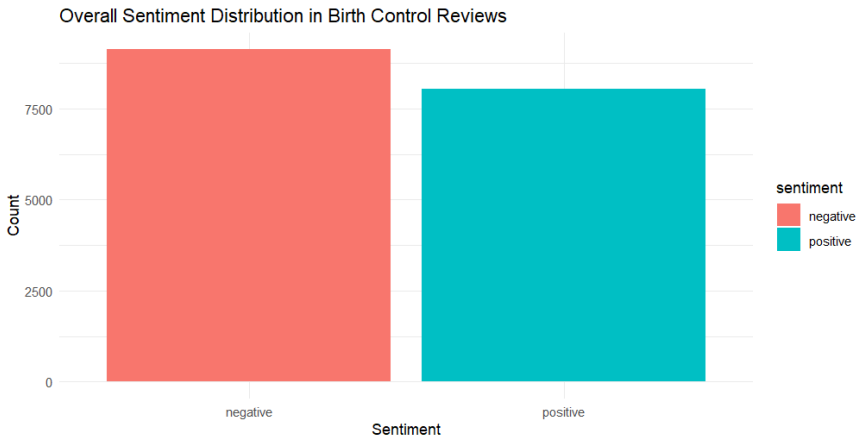
```
##
```

```
##  Spearman's rank correlation rho
##
## data:  b_sentiment$sentiment_score and b_sentiment$rating
## S = 9.8915e+10, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.314086
```

A Spearman rank correlation (rho = 0.314) ... with a highly significant p-value (< 2.2e-16) indicates a moderate but statistically significant positive association between sentiment scores and patient ratings. This suggests that patients who express more positive language in their reviews tend to give higher numerical ratings. This validates the use of sentiment analysis as a meaningful tool to quantify patient satisfaction based on unstructured text.

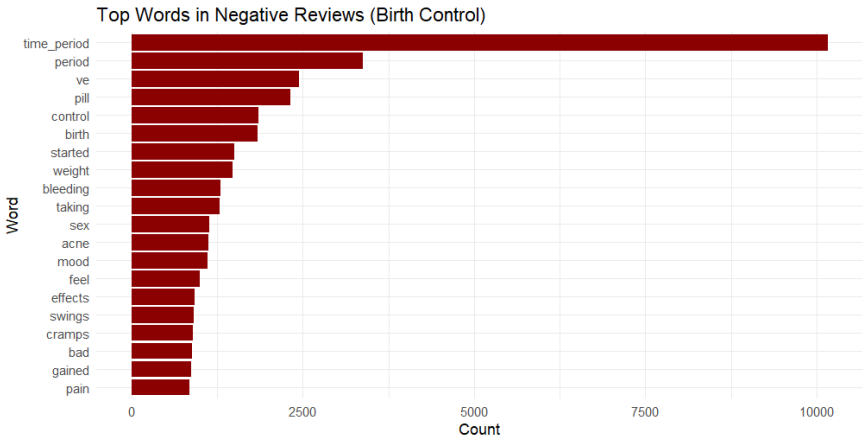
Clinical Insight: The observed correlation between positive sentiment and higher ratings could support the use of NLP techniques in post-marketing surveillance of contraceptive medications. It may help healthcare providers and regulatory bodies identify satisfaction trends and potential concerns without relying solely on structured survey tools.

Text Analysis of Negative Reviews

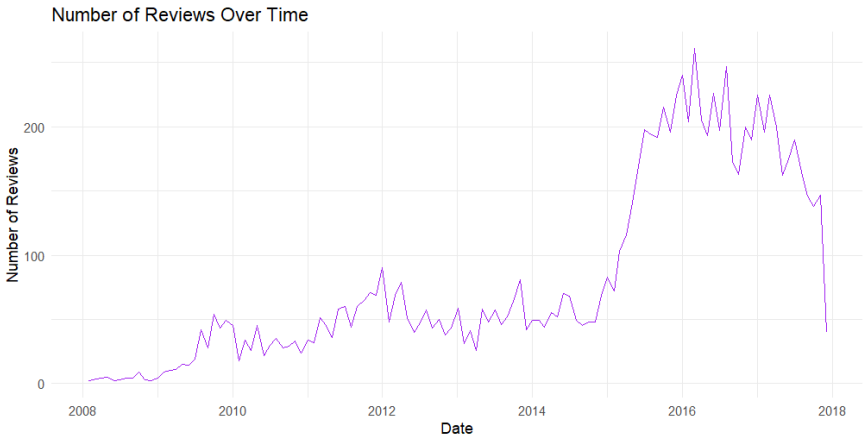


The chart shows a substantial number of reviews with negative sentiment. This may reflect frequent side effects or dissatisfaction with birth control methods. These insights could be valuable for healthcare professionals when choosing or counseling on contraceptive options.

Top Frequent Words in Negative Reviews



The most common words in negative reviews include terms like "period", "pain", and "time_period". These highlight the recurring side effects or dissatisfaction points that patients face with birth control medications.



This line plot illustrates the monthly trend in the number of patient reviews related to birth control medications. Tracking review frequency over time can reveal important patterns such as increased usage, seasonal variations, marketing influence, or changes in public awareness. Spikes or drops in review volume may correspond to events like product launches, recalls, or media coverage.

Conclousion

This analysis demonstrates the potential of using patient-generated text data for post-marketing surveillance and understanding real-world drug experiences. Birth control reviews reveal common concerns that may guide physicians in patient counseling.