Hypogammaglobulinemia in Rituximab- versus Ocrelizumab-Treated Patients with Relapsing Multiple Sclerosis

Wakeel Adekunle Kasali
wakeel.kasali@stat.ubc.ca

December 07, 2024

Abstract

Multiple sclerosis (MS) is a chronic autoimmune disease that can be managed with immune-targeting medications. This study aims to compare the impact of two immune-targeting medications, OCR and RTX, on immune proteins. It is a retrospective cohort study with roughly 350 patients conducted at the University of British Columbia Multiple Sclerosis Clinic between 2017 and 2024. Side-by-side boxplots and contingency tables are recommended to summarize and compare demographic variables between OCR and RTX. In addition, to compare medications by patients, spaghetti plots are recommended to visualize trends in lab measured immune protein levels and Kaplan-Meier survival curves are recommended to compare times to lab value deficiencies. Logistic regression is recommended to compare probability of developing lab value deficiencies within a certain time on different medications. A Cox proportional hazards model is recommended to model time to lab value deficiencies.

1 Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease in which the immune system attacks the brain and spinal cord, leading to progressive neurological symptoms such as motor dysfunction, sensory impairments, and cognitive decline. The relapsing-remitting form of MS (RMS) represents a key clinical subtype often managed with immune-targeting medications. Rituximab (RTX), covered by BC Pharmacare, and Ocrelizumab (OCR), not covered, are two most used medications with distinct access and cost implications. However, the specific immunological effects of these medications on some key immune components remain under-researched. This study addresses this gap by investigating how these medications influence the deficiency in the immune system function. The information collected includes immunoglobulin levels (IgG, IgA, IgM), lymphocyte counts, neutrophil counts, and follow-up laboratory values. It is hypothesized that patients with RMS treated with OCR display distinct patterns of immunoglobulin deficiencies compared to those treated with RTX, with potential implications for immune competency. The following sections describe the dataset, formulate the key statistical questions, and outline the recommended methods for analysis

2 Data Description

A retrospective cohort study is conducted at the University of British Columbia Multiple Sclerosis Clinic between 2017 and 2024. The study includes roughly 350 patients with a confirmed diagnosis of relapsing-remitting multiple sclerosis (RMS) who are treated with either OCR or RTX for at least one year. As previously mentioned, the assignments of the two treatments are based on the patients' health insurance coverage and therefore are not randomized. Before treatment, patients' immune protein levels, including

IgG, IgM, IgA, neutrophils, and lymphocytes, are recorded. During the follow-ups after treatment, the same lab values are monitored and documented. The dataset includes the study ID, age, sex, income quantile, treatment type (RTX or OCR), start date of the treatment, treatment duration (calculated as the difference between the latest lab visit and the start date) and information about each lab visit: visit date, measured lab values. Each lab value has a known lower threshold indicating deficiency but the dataset does not have deficiency indicator variables for all lab visits. Due to OCR being introduced a few years earlier, the average treatment length for OCR patients is 60 months compared to an average of 40 months for RTX patients.

There are missing pre-treatment lab values in both the RTX and OCR groups, with a higher frequency observed in the OCR group. During the pre-treatment lab data collection, it is assumed that the immune system might be less affected, so the pre-treatment values are not consistently recorded. In the post-treatment lab data, the maximum number of post-treatment lab visits is seven and the intervals between two consecutive lab visits are irregular. For instance, a patient may complete the first (pre-treatment) and second lab visits and then drop out of subsequent follow-ups.

3 Statistical questions

The main statistical questions for the study are as follows:

- How does the incidence of lab value deficiencies (e.g., IgG, IgM, IgA, neutrophils, and lymphocytes) differ between the two treatments within a fixed number of years (e.g., 2 years) on treatment? What are the associated risk factors?
- How can the time to lab value deficiencies (defined as the time from start of treatment to the first time immunoglobulin levels fall below deficiency thresholds) be modeled?

4 Proposed Statistical Methods

The statistical methods recommended to address the two statistical questions are described as follows:

4.1 Exploratory Data Analysis

Side-by-side boxplots and contingency tables are recommended to summarize and highlight potential associations between the demographic variables and treatment assignment in the dataset. Figure 1 illustrates the age distribution of patients for each of the two treatments.

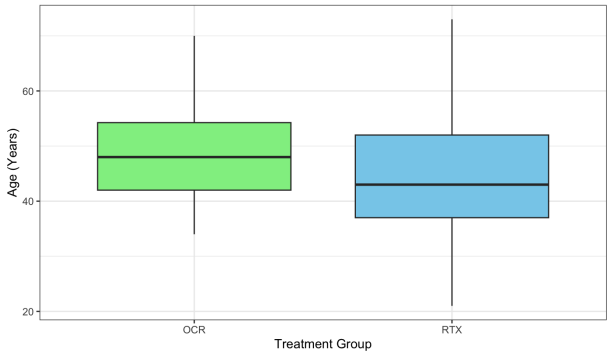

Figure 1: Distribution of age for OCR and RTX from provided dataset.

Income quantiles derived from postal codes serve as a proxy for socioeconomic status (SES). SES provides insights into healthcare access, which may influence the treatment assignment. Table 1 summarizes

treatment assignment across income quantiles. Socioeconomic status, as represented by income quantiles, can significantly impact immune responses and thus guide understanding about how treatment effects may vary across income groups. For instance, as in Table 1, OCR which is not covered by insurance may appear to have more participants in higher income quantiles (e.g., quantiles 4 and 5), whereas RTX has a larger proportion of participants in the lower income quantiles (e.g., quantiles 1 and 2). This disparity can highlight potential biases in treatment assignment.

Table 1: Contingency table of income quantiles versus treatment assignment from provided dataset.

Income Quantile	OCR	RTX
1	19	40
2	29	35
3	33	36
4	40	33
5	52	28

Spaghetti plots are recommended to visualize the trend in the lab values. Immunoglobulin levels might show both stability and sudden changes across time (in months) since treatment started. Some patients might exhibit consistent trends, while others could display irregular fluctuations. At specific lab visits, the Immunoglobulin values might be similar or close across individuals. Figure 2 provides intuition that could help to understand these variations in Immunoglobulin G levels. From this plot, patients with consistently high IgG levels seems to have shorter treatment times. A similar plot is recommended for other antibodies to explore potential patterns and trends across lab visits.

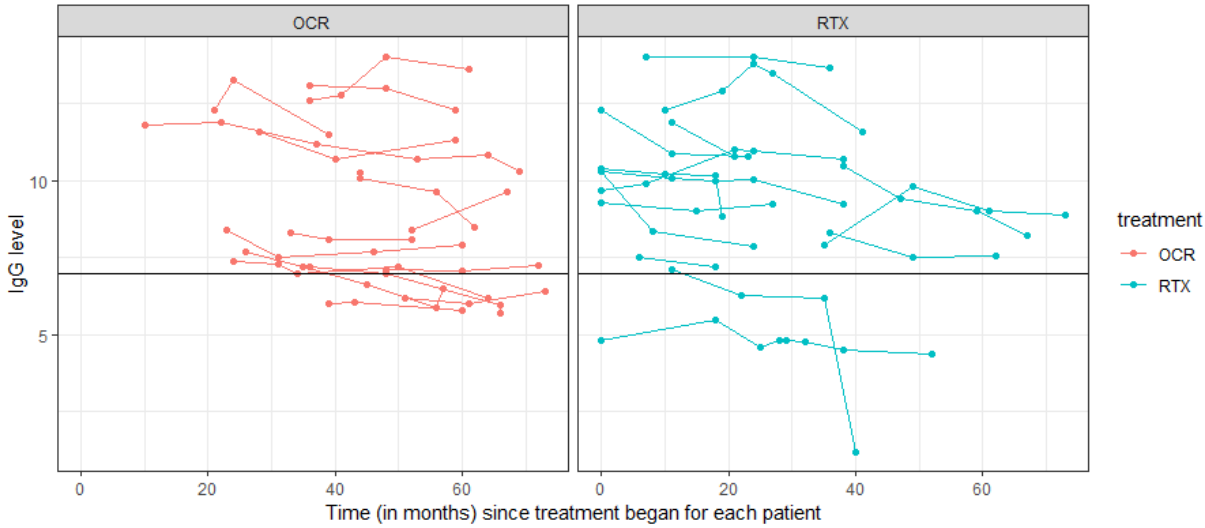

Figure 2: Longitudinal trends in Immunoglobulin G (IgG) levels across lab visits for 15 random OCR patients and 15 random RTX patients based on the study dataset. The trends not starting at 0 are from patients whose pre-treatment IgG levels are missing.

To visualize time to lab value deficiency, Kaplan-Meier survival curves are recommended. In comparing treatments, survival probabilities for avoiding antibody deficiency might be higher with one medication than with other, particularly after consistent use over time. This suggests that one therapy could delay deficiency more effectively in the long term. In the initial months of treatment, there might be little to no noticeable difference between the two medications in preventing deficiency. However, as time progresses, the survival probabilities could diverge, possibly due to fewer participants remaining in the study. Figure 3 illustrates these trends using IgG deficiency as an example.

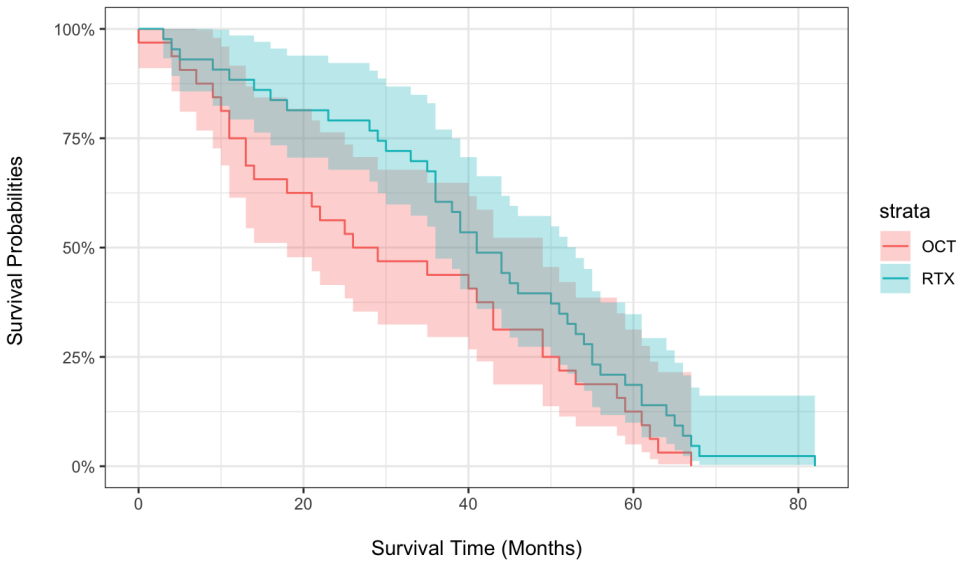

Figure 3: Kaplan-Meier survival curves comparing time to IgG deficiency between OCT and RTX treatment groups based on the study dataset

4.2 Analysis of Lab Value Deficiency Incidence

To compare probability of having lab value deficiency after a fixed amount of time of treatment initiation (e.g., 2 years) between RTX and OCR, we recommend using a logistic regression model.

Logistic regression Logistic regression for a binary response variable assumes the log-odds of a binary outcome is a linear combination of one or more explanatory variables. In equation form, a logistic regression model with explanatory variables $\mathbf{x} = (x_1, ..., x_m)$ satisfies

$$\log\left(\frac{P(Y=1|\mathbf{x})}{1-P(Y=1|\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m \tag{4.1}$$

where Y is a binary indicator of lab value deficiency within a specified time period (such as 6 months or a year from start of treatment), $x_1, \ldots, x_m$ are explanatory variables such as age, gender and treatment, and $\beta_0, \ldots, \beta_m$ are the coefficients of the model. The coefficient β_j represents the change in log-odds of the probability of lab value deficiency within a specified time period when the corresponding continuous variable X_j increases by 1 unit while holding all other variables constant. For categorical variables, the β_j coefficients represent the change in the log-odds of the probability of lab value deficiency within a specified time period for each category relative to a user-defined reference level with all other variables held constant.

Variable selection To select which variables to include, we recommend using the Akaike Information Criterion (AIC), which chooses explanatory variables which predict well without overtting. As the number of explanatory variables is small (roughly 5) we can go through the list of possible variable subsets exhaustively and choose the subset with the smallest AIC.

4.3 Time to Deficiency Analysis

To model the time to lab value deficiency, we recommend using a Cox proportional hazards model.

The Cox proportional hazards model is a popular survival analysis model used to examine the effect of explanatory variables on the time to an event (in this case the time to reach lab value deficiency). In particular, the hazard function h_i of patient i with explanatory variables $\mathbf{x}_i$ at time t is modelled as

$$h_i(t) = h_0(t)\exp(\mathbf{x}_i\beta) \tag{4.2}$$

where h_0 is the baseline hazard function and β is the coefficient vector of effects. Intuitively, if $h_i(t) > h_j(t)$ then patient i is at higher risk of lab value deficiency than patient j at time t given that both their lab values have not fall below the deficiency threshold.

The Cox model has several benefits. First, the Cox model does not assume a parametric form for the baseline distribution for survival times which makes it quite flexible. Second, it can handle right-censored and interval-censored survival times. Here right-censored means the time-to-deficiency falls within the interval (t, ∞) where t denotes the time from start of treatment to the date of the latest lab result and interval-censored means the time-to-deficiency falls within the interval (a, b) where a, b denotes the times from start of treatment to the two lab visits between which the deficiency occurs. For example, if lab 2, which happened 12 months after start of treatment, indicates no deficiencies and lab 3, which happened 24 months after start of treatment, indicates a deficiency then the time-to-deficiency is in the interval (12, 24) months after start of treatment. To handle interval censoring, we recommend using the `R` package `icenReg` as demonstrated in the Appendix. In addition, we recommend fixing the maximum follow-up time to be that of the newer treatment (i.e., RTX) and censoring any survival time longer than that to adjust for the imbalance in censoring time between the two treatments as well as using variables previously chosen in the logistic regression model. Note, however, that the estimation of hazard ratios become unreliable if too many deficiency times are right-censored or if censored intervals are too wide.

5 Conclusions and Recommendations

In summary, we recommend side-by-side boxplots, contingency tables to summarize the demographic variables of the dataset, spaghetti plots to visualize trends in lab measured immune protein levels and Kaplan-Meier survival curves to compare times to lab value deficiencies for the two medications. Logistic regression is recommended to compare probability of developing lab value deficiencies within a fixed time on different medications and variables are chosen based on AIC. A Cox proportional hazards model is recommended to model the right and interval censored times to lab value deficiencies.

6 Further reading

- For an overview of exploratory data analysis techniques: https://asda.stat.ubc.ca/Workshops/ASDaStatsResources/EDA_Tutorial.html
- For further reading on logistic regression models: Applied Logistic Regression, 2nd ed, Hosmer and Lemeshow (2000), Wiley.
- For further reading on time-to-event analysis: Statistical Models and Methods for Lifetime Data, 2nd ed, Lawless (2003), Wiley.

Appendix

The following shows a sample R implementation of the analysis in Section 4 on the provided dataset.

```
library(ggplot2)
library(ggcorrplot)
library(MASS)
library(dplyr)
```

```r
# install.packages(c("survival", "survminer", "icenReg"))
# packages for time-to-event analysis
library(survival)
library(survminer)


# example uses IgG but analysis can also be applied to other lab values
# read data
dataMS = read.csv("client_data.csv", header = T)
# remove patients with no treatment label or no after treatment lab values
data_filtered = dataMS %>% filter(Length.of.Tx > 12, Treatment != "", !is.na(IgG.2))
data_filtered[data_filtered == ""] = NA # set empty cells to NA
for (j in 2:8) { # set labs with date but no lab value to NA
  month_col = paste0("Months.",j)
  lab_col = paste0("IgG.",j)
  for (i in 1:dim(data_filtered)[1]) {
    if (is.na(data_filtered[lab_col][i,1])) {
      data_filtered[month_col][i,1] = NA
    }
  }
}
data_filtered = data_filtered[-62,] # remove obs with conflicting timeline
# rename treatments
data_filtered$Treatment[data_filtered$Treatment=="O"] = "OCR"
data_filtered$Treatment[data_filtered$Treatment=="R"] = "RTX"

# Dataset columns are in the form
# Study.ID | Age | Sex | Income.Quantile | Treatment | Lab1.Date | Lab1.IgG | Lab2.Date |
# Here each row contains 1 patient's info and all lab values
# To do time-to-event analysis, we need to obtain the censored interval of each patient


## Reformat data for analysis and spaghetti plot
id = c()
treatment = c()
months = c()
IgG = c()
IgG.Low = c() # IgG low indicator for each recorded lab
IgG.Low2 = c() # IgG low indicator for Cox model
left.interval = c() # left side of censored interval
right.interval = c() # right side of censored interval


# function to record and check if current patient-lab pair shows deficiency
deficiency_check = function(month_col, lab_col, i){
  if ((!is.na(data_filtered[month_col][i,1])) || (month_col == "Months.1")){
    id <<- c(id,data_filtered$Study.ID[i])
    treatment <<- c(treatment,data_filtered$Treatment[i])
    months <<- c(months,data_filtered[month_col][i,1])
    IgG <<- c(IgG,data_filtered[lab_col][i,1])
    IgG.Low <<- c(IgG.Low,data_filtered[lab_col][i,1]<7) # 7 is the IgG deficiency thresho
  }
}
```

```
# This code loops over all patients and their lab visits to record IgG deficiencies
for (i in 1:dim(data_filtered)[1]) { # loop over all patients
  IgG.Low_temp = 0
  left.interval_temp = 0
  right.interval_temp = 0
  for (j in 1:8) { # loop over all lab visits of a patient
    deficiency_check(paste0("Months.",j), paste0("IgG.",j), i)
    if (j>1){
      if (!is.na(data_filtered$IgG.1[i]) & (IgG.Low[length(IgG.Low)-1]==0) &
          (IgG.Low[length(IgG.Low)]==1)){
        IgG.Low[length(IgG.Low)] = 3 # interval censored
        IgG.Low_temp = 3
        left.interval_temp = months[length(months)-1]
        right.interval_temp = months[length(months)]
      } else if ((j==2) & is.na(data_filtered$IgG.1[i]) & (IgG.Low[length(IgG.Low)]==1)) {
        IgG.Low[length(IgG.Low)] = 3
        IgG.Low_temp = 3
        left.interval_temp = months[length(months)-1]
        right.interval_temp = months[length(months)]
      }
    }
  }
  IgG.Low2 = c(IgG.Low2,IgG.Low_temp)
  if(right.interval_temp == 0){ # no deficiency detected
    left.interval_temp = months[length(months)]
    right.interval_temp = Inf
  }
  left.interval = c(left.interval,left.interval_temp)
  right.interval = c(right.interval,right.interval_temp)
}


# data for spaghetti plot (each row is a patient-lab visit pair)
data_spaghetti = data.frame(id=id,treatment=treatment,months=months,IgG=IgG)
# spaghetti plot data view
#   id treatment months   IgG
# 1  8       RTX      0 10.00
# 2  8       RTX      7  9.80
# 3  8       RTX     12  9.20
# 4  8       RTX     25  8.56
# 5  9       RTX      0  9.40
# 6  9       RTX     47  7.84

# data reformatted for analysis (get rid of Lab value, Lab date columns and
# compute censored intervals)
data_reformatted = data.frame(age=data_filtered$Age, sex=data_filtered$Sex,
        income=data_filtered$Income.Quantile, treatment=data_filtered$Treatment,
        left.interval=left.interval, right.interval=right.interval, IgG.Low=IgG.Low2,
        IgG.Low.2y=(right.interval<24)) # last col indicate deficiency within 2 years

# Reformatted dataset view
# IgG.Low: 0=right-censored, 3=interval-censored
```

```
#    age sex income treatment left.interval right.interval IgG.Low IgG.Low.2y
# 1   40   F      3       RTX            25            Inf       0      FALSE
# 2   51   M      3       RTX            59             62       3      FALSE
# 3   45   F      1       RTX            22            Inf       0      FALSE
# 4   48   M      2       RTX            30            Inf       0      FALSE
# 5   42   M      1       RTX            66            Inf       0      FALSE
# 6   40   F      2       RTX            23            Inf       0      FALSE
```

```
## Spaghetti plot
idx <- c()
for (i in data_filtered$Study.ID) { # remove patients with only 1 lab visit
  if (sum(!is.na(data_spaghetti$IgG[data_spaghetti$id == i])) > 1){
    idx <- c(idx,i)
  }
}
data_spaghetti = data_spaghetti[data_spaghetti$id %in% idx,]

# extract 15 patients from each treatment for spaghetti plot
set.seed(123)
id_OCR = sample(unique(data_spaghetti[data_spaghetti$treatment=="OCR",]$id),15)
id_RTX = sample(unique(data_spaghetti[data_spaghetti$treatment=="RTX",]$id),15)
idx = data_spaghetti$id %in% c(id_OCR,id_RTX)
p = ggplot(data = data_spaghetti[idx,], aes(x = months, y = IgG, group = id)) +
  geom_line(aes(color=treatment)) + geom_point(aes(color=treatment)) +
  geom_hline(yintercept=7) +
  xlab("Time (in months) since treatment began for each patient") +
  ylab("IgG level") + facet_grid(. ~ treatment)
p
```



```
## Logistic regression
# assume that age and sex are used along with treatment
res.logistic = glm(IgG.Low.2y ~ age + sex + treatment,
                   data = data_reformatted, family = binomial)
summary(res.logistic)

# Call:
# glm(formula = IgG.Low.2y ~ age + sex + treatment, family = binomial,
#       data = data_reformatted)
#
# Coefficients:
#             Estimate  Std. Error  z value  Pr(>|z|)
# (Intercept) -5.41633     1.56838   -3.453  0.000553 ***
# age          0.04746     0.02957    1.605  0.108443
```

```
# sexM         -0.74716    0.78950  -0.946 0.343960
# treatmentR    0.09785    0.58301   0.168 0.866716
# ---
# Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1          1
#
# (Dispersion parameter for binomial family taken to be 1)
#
# Null deviance: 110.75  on 344  degrees of freedom
# Residual deviance: 107.21  on 341  degrees of freedom
# AIC: 115.21
#
# Number of Fisher Scoring iterations: 6


## Time-to-event analysis
# package for interval censored data
library (icenReg)

# assume that age and sex are used along with treatment
res.cox = ic_sp(Surv(left.interval, right.interval, IgG.Low,type = "interval") ~ age + sex
                data = data_reformatted, bs_samples = 100)
summary(res.cox)

# Model:  Cox PH
# Dependency structure assumed: Independence
# Baseline:  semi-parametric
# Call: ic_sp(formula = Surv(start, stop, IgG.Low, type = "interval") ~
#               age + sex + treatment, data = data_Cox, bs_samples = 100)
#
#            Estimate Exp(Est) Std.Error z-value       p
# age         0.03954   1.0400   0.01707  2.3160 0.02058
# sexM       -0.58090   0.5594   0.40140 -1.4470 0.14790
# treatmentR -0.10690   0.8986   0.34610 -0.3089 0.75740
#
# final llk =  -142.6641
# Iterations =  15
# Bootstrap Samples =  100

# coefficient interpretation: a positive sign means
# that the hazard (risk of deficiency) is higher and vice versa
```