

# **HIV Treatment: The Short-Term and Long-Term Effects of HIV Therapy**

Nikita Jayaprakash

## **Introduction**

While there is existing research on the effects of human immunodeficiency virus (HIV) treatment of the survival among patients with advanced HIV, there are still gaps in the knowledge relating to prolonged clinical benefit and the effects of the stage of the disease and the prior HIV therapy on the efficacy of certain HIV treatments. It is known that Drug A improves survival among patients with advanced HIV and slows the progression of disease in patients with little to no symptoms. We hypothesize that there exists prolonged clinical effect for patients with less advanced patients. Research shows that effective treatment of HIV disease will likely require combination therapy, but we hypothesize that there is a difference in responses to HIV therapy according to the stage of the disease and extend of prior HIV drug exposure.

To address these issues, this analysis uses data from a randomized clinical trial that compared monotherapies with Drug A alone or Drug B along with combination therapies – such as Drug A and Drug B together or Drug A and Drug C together – in adults infected with HIV. The following questions are explored:

- Do responses to HIV therapy vary according to the stage of the disease?
- Do responses to HIV therapy vary according to the extent of prior HIV drug exposure?
- Is there a prolonged clinical benefit from drug intervention for patients with less advanced HIV disease?

It should be noted that not all initial patients were treated for the entirety of the trial, so the causal conclusions of this analysis are conditional on the assumption of no confounding based on this censoring effect. CD4 cells are white blood cells that fight infection, and CD8 are suppressor cells that complete the immune response. A normal CD4/CD8 ratio is 2.0 that is higher than 2 indicates a strong immune system while a ratio is less than 1 possibly indicates HIV, AIDS if the CD4 count is less than 200/mm or other serious problems. In addition, we will use the ratio of CD4 to CD8 cells and the difference in CD4 cell count across the study to assess stage of HIV disease. CD4 cells are helper cells, which lead the attack against

infections, and CD8 are suppressor cells, which complete the immune response, so they are potential confounders.

This analysis determined that following conclusions. Response to HIV therapy do vary according to stage of the disease, and patients with less advanced stages of the disease have a more positive response to the HIV therapy. Responses to HIV therapy also vary according to the extent of prior HIV drug exposure to a statistical significant degree. There is positive prolonged clinical benefit from drug intervention for patients with less advanced HIV disease based on the different of their CD4 cell count.

## Data

The original data contained contains 2139 observations from a randomized clinical trial with 27 variables, relating to various cell counts of patients throughout the trial, details about their treatment and their symptoms, their demographic information and information about previous therapy and relevant activities like homosexual activity or a history of intravenous drug use. There is one extreme outlier who had a CD8 count at baseline of 40, who was excluded from the analysis. 776 patients were off-treatment before  $96 \pm 5$  weeks, so these patients were dropped from the analysis In addition, of the remaining 1362 patients, 287 patients did not have a recorded CD4 cell count at 96 weeks, so these patients are excluded from the analysis of the third research question because their response to HIV therapy is measured at 96 weeks to determine prolonged clinical benefit. The sample for the third research question is 1075 total patients. The missing CD4 cell counts at 96 weeks are the only missing values of this dataset. Once again, any causal conclusions of this study assume that the censoring effect is not confounding between patients who continued their treatment and patients who didn't, and that the CD4 cell count is not confounded with the CD8 cell count at 96 weeks. For this reason, it would be helpful if we had additional variables that tracked CD8 cell count at 96 weeks for all patients and that explained why patients stopped treatment in order to assess the accuracy of these assumptions and adjust for any confounding, if applicable.

### The 20-Week Model

The first two research questions will use the same regression model that models patients from baseline through 20 weeks. For clarity, this model will be referenced at the “20-week model”. The response variable must determine responses to HIV therapy. The ratio of the CD4 to CD8 at 20 weeks will be used as the response variable for the first two research questions because it indicates a strong or comprised immune system. Similarly, for the first research question, the covariate of main interest – which must determine the stage of the disease at the beginning of the study – will be the ratio CD4/CD8 at baseline. The distribution of CD4/CD8 at 20 weeks and at baseline are slightly right-skewed (see Figure 1).

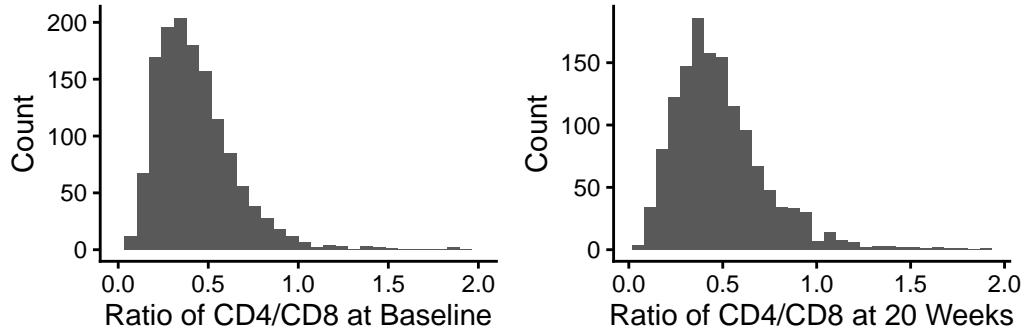


Figure 1: There is one outlier with a baseline CD4/CD8 ratio of 10 – CD4 count of 400 and CD8 count of 40 – that is included in the model but excluded from the baseline plot.

For the second research question, the extent of prior drug exposure will be represented by three variables: whether the patient had HIV therapy other than Drug A prior to initiation of study treatment, whether the patient used Drug A in the 30 days prior to treatment initiation and the number of days of previously received HIV therapy for each patient. Of the 1363 patients, only 27 had HIV therapy other than Drug A prior to initiation of study treatment, and 765 had Drug A use in the 30 days prior to treatment initiation (see Figure 2).

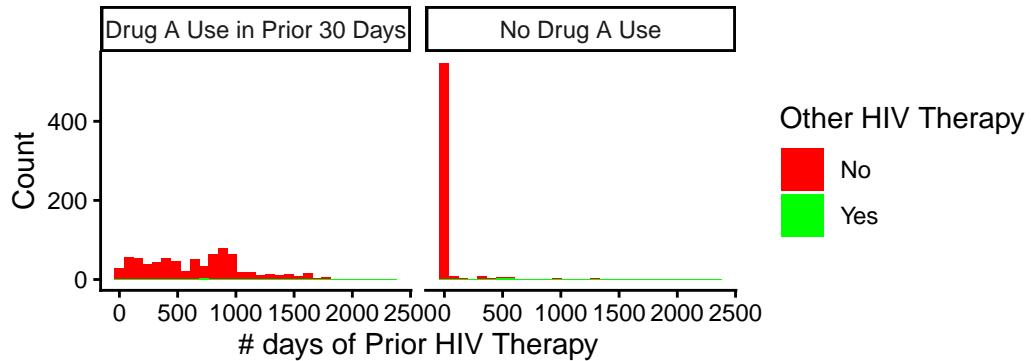


Figure 2: Very few patients have received other HIV therapy aside from previous Drug A usage.

### The 96-Week Model

For third research question, the response variable must determine prolonged clinical benefit. Since the full study recorded values up until 96 weeks, this analysis will consider the 96-week vitals as prolonged results, and this model will be referenced as the “96-week model” for clarity. However, this data only includes the CD4 count at 96 weeks; due to this limitation, for the third research question, the analysis will use difference of CD4 cell count from 96 weeks to

baseline in order to determine the prolonged clinical benefit of HIV therapy. Since CD4 count of 500-1500 per millimeter is healthy and of 200-500 signals a weak immune system, this will be encoded as a binary variable that classified patients with a CD4 cell count over 500 as less advanced HIV status. We are assuming there is no confounding between the CD8 cell count at 96 weeks. The distributions for CD4 cell counts at 96 weeks and at baseline are relatively symmetric with no visual outliers (see Figure 3).

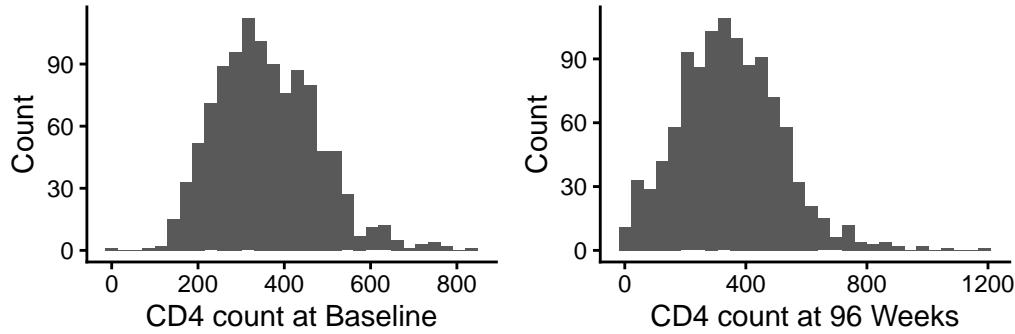


Figure 3: The CD4 count have a larger range for patients at 96 weeks than at baseline.

## Included Variables

The following variables should be included in all fitted models in order to account for additional variation in the response variables of responses to HIV therapy or prolonged clinical benefit:

- demographic variables including race (whether or not they are white), gender (female vs. male), age in years and weight in kilograms at the start of the treatment
- relevant activity such as homosexual activity or history of intravenous drug use
- symptoms, such as if they were symptomatic and if they have hemophilia
- treatment type, which could be Drug A alone, Drugs A and B, Drugs A and C and Drug B alone
- unfavorable events, such as number of days until the first occurrence of an unfavorable event, whether or not they have an unfavorable event by the time point and the interaction between the two

For the demographic variables, there are no visible trends between the ratio of CD4 to CD8 at 20 weeks, but we will include because of the conceptual relevance; in addition, there are some outliers with a higher ratio across the race and gender demographics (Figure 4). There are also several outliers across all symptom- and activity-related variables, but nothing is concerning enough to exclude (Figure 5). Patients who were censored tended to have unfavorable events near the end of their treatment, so there is possible confounding because of the censoring effect (Figure 6). This is a limitation of this study because we do not have more information about

these censored patients. It would be helpful to have more information about why they dropped out of treatment.

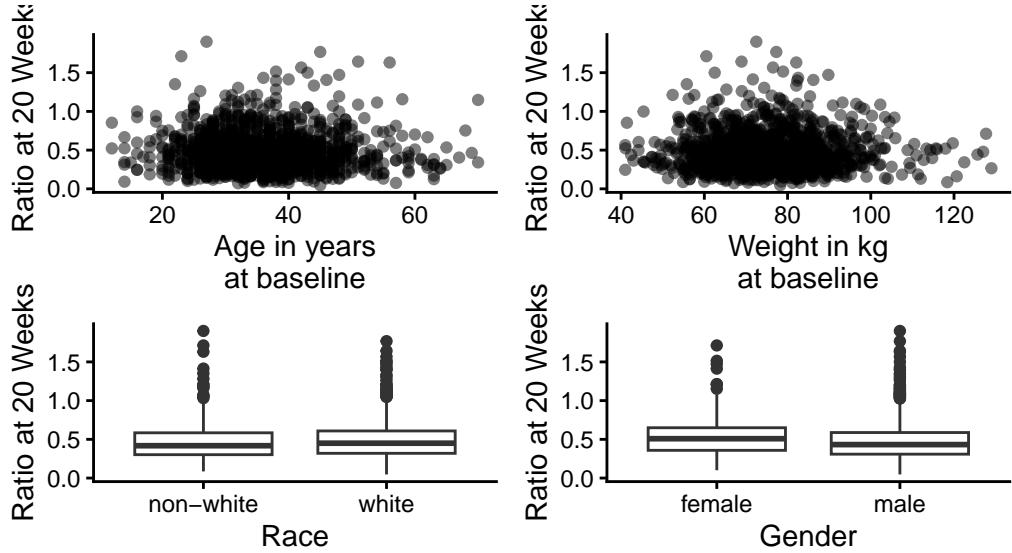


Figure 4: There are no visible trends between the ratio CD4/CD8 at 20 weeks and the demographic variables.

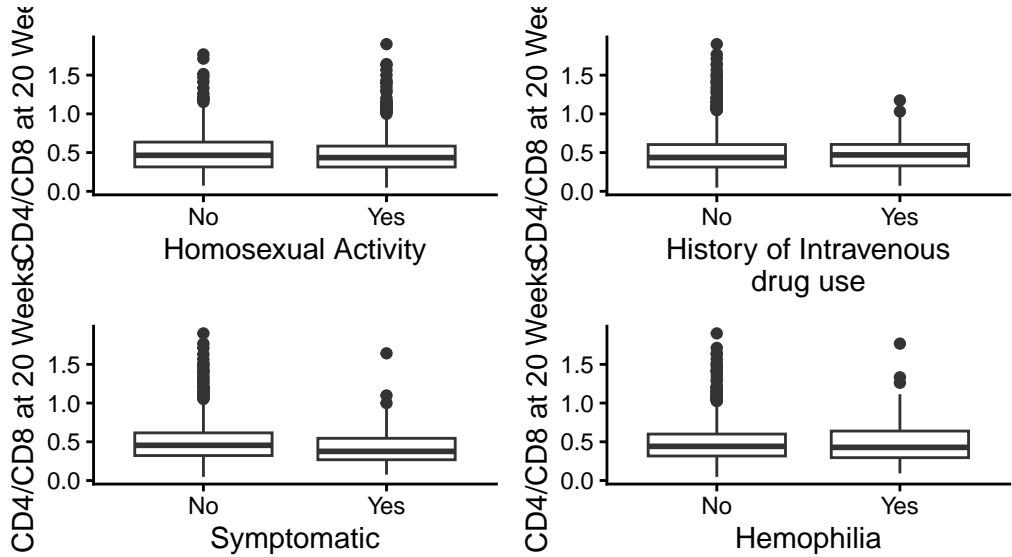


Figure 5: There are no visible trends between the ratio CD4/CD8 at 20 weeks and the variables about relevant activity and symptoms. However, there are distinct outliers across all levels of all variables.

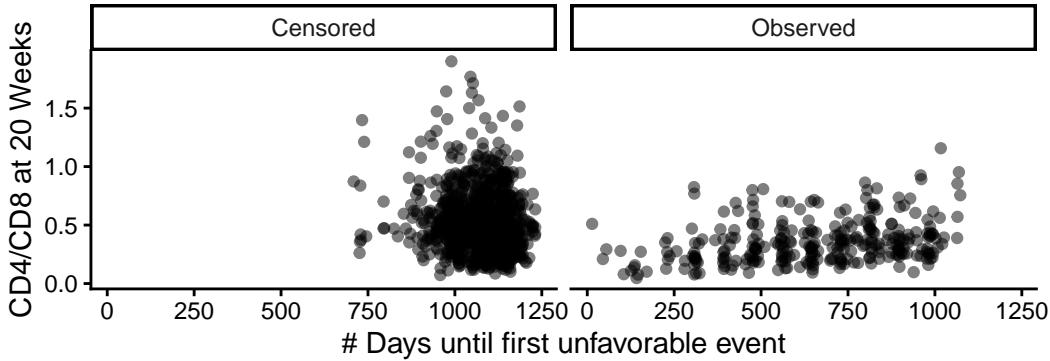


Figure 6: Patients who had an observed unfavorable event had a varied range of how soon this event occurred in the clinical trial.

In addition, the fitted model for the first and third research question will also include three variables used to determine the extent of prior drug exposure in the second model (i.e. whether the patient had HIV therapy other than Drug A prior to initiation of study treatment, whether the patient used Drug A in the 30 days prior to treatment initiation and the number of days of previously received HIV therapy for each patient). The fitted model for the third research question will also include the CD4 count at 20 weeks in order to account for variation in prolonged clinical benefit from the short-term effects.

## Excluded Variables

The patient's ID number is an identifier of individual patients and is not a relevant indicators of any of the response variables. This analysis also excludes Karnofsky score at baseline, which is a diagnosis that measures functional impairment and is a manmade score that can be biased. There are two categorical variables that determine HIV therapy history prior to study and HIV therapy history stratification respectively; history of HIV therapy is more accurately represented by the number of days of previously received HIV therapy, so these two variables are excluded. Similarly, there is an indicator variable that determined whether or not a patient received any form of treatment, but the treatment arm variable gives more specific information (i.e. which treatment they got) and is more relevant in the model. There is an indicator variable that determines if a patient has Drug A use prior to treatment initiation; all filtered patients had some form of Drug A usage, so this is excluded.

## Methods

In this report, we fit two linear regression models. The 20-week model will regress the ratio CD4/CD8 cell counts at 20 weeks on the ratio CD4/CD8 at baseline and the additional

predictors of age, weight, race, gender, hemophilia, indicator of homosexual activity, indicator of previous intravenous drug use, whether the patient had HIV therapy other than Drug A prior to initiation of study treatment, whether the patient used Drug A in the 30 days prior to treatment initiation, the number of days of previously received HIV therapy for each patient, if they were symptomatic, treatment arm, number of days until the first occurrence of an unfavorable event, whether or not they have an unfavorable event by the time point and the interaction between the two. This model will be used to analyze the first two research questions. This analysis will create a 95% confidence interval on the coefficient for the ratio of CD4 to CD8 at baseline in order to determine if there is a causal relationship between responses to HIV therapy and the stage of the disease. This analysis will also conduct one partial F-test at 5% significance level. The reduced model for the partial F-test that answers the second research question will exclude the three variables that determine the extent of prior drug exposure, as previously mentioned. This will determine if there is a causal relationship between responses to HIV therapy and the extent of prior HIV drug exposure.

The 96-week model will regress the CD4 cell count at 96 weeks on the binary indicator of whether a patient has less advanced stage of HIV – determined by a cutoff of their CD4 cell count at baseline as explained above – and all the same additional predictors as the first model with the inclusion of the CD4 cell count at 20 weeks. This model will be used to analyze the third research question. This analysis will conduct a 95% confidence interval of the coefficient of the binary indicator of less advanced stage of HIV to test if there is a prolonged clinical benefit from drug intervention for patients with less advanced HIV disease.

The assumptions for using a linear regression model are believed to be satisfied. We assume that there is no confounding of the censoring effect on patients who were treated through 96 weeks and those who dropped out. Since this is a randomized experiment through a clinical trial, the  $Y|X$  are mutually independent by design. The variance of  $Y|X$  is also likely constant with mean 0 because the variance of bodily functions and diseases tend to have constant variance across various patients. Data related to bodily functions tend to be normally distributed. Since this analysis utilized partial F-tests and confidence intervals, this is not a crucial assumption and the large sample size allows the assumption of approximate normality.

These validity of the assumptions and the general accuracy of the model will be assessed after fitting the linear regression through the use of diagnostic plots such as a residual v. fitted plot and a Q-Q plot. The diagnostic plots for the first model (see Figure 7) and the second model (see Figure 8) show that there are no signs of nonconstant variance aside from a few outliers. The QQ plots from diagnostic plots for the two models (Figure 7 and Figure 8) show that the normality assumption holds since the data is largely normally distribution with some slight deviation at the tails. Overall, there is no concern against the assumptions of this dataset.

## Results

### The 20-week Model

The first fitted model is:

$$Ratio_{20weeks} = \beta_0 + \beta_1 * Ratio_{baseline} + \beta_2 * Age + \beta_3 * Weight + \beta_4 * Race_{Non-White} + \beta_5 * Gender_{Male} + \beta_6 * Hemophilia + \beta_7 * Homosexual Activity + \beta_8 * Intravenous Drug + \beta_9 * Prior HIV Therapy_{not drug A} + \beta_{10} * Prior Drug A_{30days} + \beta_{11} * Days of Prior HIV Therapy + \beta_{12} * Symptomatic + \beta_{13} * Treatment Arm_{DrugsAand} + \beta_{14} * Treatment Arm_{DrugsAandC} + \beta_{15} * Treatment Arm_{DrugsBalone} + \beta_{16} * Unfavorable Event_{observed} + \beta_{17} * Unfavorable Event_{#days} + \beta_{18} * Unfavorable Event_{observed : Unfavorable Event_{#days}} + \epsilon$$

Using the 20-week model, the 95% confidence interval for  $\beta_1$  – the coefficient for the ratio of CD4 to CD8 at baseline – in our linear model, is [0.8367507, 0.9102958]. Since zero is not included in this interval, we reject the null hypothesis at a 5% significance level that there is no significant relationship between responses to HIV therapy and the stage of the disease, assuming all other predictors in the model are held constant. For  $\beta_1 = 0.8735$ , people with less advanced HIV (i.e. a higher ratio of CD4 to CD8) have a greater response to HIV therapy based on their ratio of CD4 to CD8 at 20 weeks. Using the 20-week model, at a 2.5% significance level, we also reject the null hypothesis (i.e.  $H_0 : \beta_9 = \beta_{10} = \beta_{11} = 0$ ) that there is no causal relationship between responses to HIV therapy and extent of prior HIV drug exposure ( $F(3, 1343) = 12.17$ ,  $p < 0.001$ ). Note that the error term  $\epsilon$  is assumed to be independent and identically distributed with mean 0 and constant variance.

### The 96-week Model

The second fitted model is:

$$Diff_{96weeks} = \alpha_0 + \alpha_1 * Less\ Advanced_{Baseline} + \alpha_2 * Age + \alpha_3 * Weight + \alpha_4 * Race_{Non-White} + \alpha_5 * Gender_{Male} + \alpha_6 * Hemophilia + \alpha_7 * Homosexual Activity + \alpha_8 * Intravenous Drug + \alpha_9 * Prior HIV Therapy_{not drug A} + \alpha_{10} * Prior Drug A_{30days} + \alpha_{11} * Days of Prior HIV Therapy + \alpha_{12} * Symptomatic + \alpha_{13} * Treatment Arm_{DrugsAand} + \alpha_{14} * Treatment Arm_{DrugsAandC} + \alpha_{15} * Treatment Arm_{DrugsBalone} + \alpha_{16} * Unfavorable Event_{observed} + \alpha_{17} * Unfavorable Event_{#days} + \alpha_{18} * Unfavorable Event_{observed : Unfavorable Event_{#days}} + \alpha_{19} * CD4\ Count_{20weeks} + \epsilon$$

Using the 96-week model, the 95% confidence interval for  $\alpha_1$  – the coefficient for the indicator of less advanced patients based on CD4 cell count at baseline – in our linear model, is [-135.0561, -88.33476]. Since zero is not included in this interval, we reject the null hypothesis at a 5% significance level that there is no significant relationship between prolonged clinical benefit from drug intervention and whether a patient has less advanced HIV disease, assuming all other predictors in the model are held constant. Because  $\alpha_1 = -111.7$  is negative for less

advanced patients, this means that there is a positive prolonged clinical benefit in HIV therapy for patients with less advanced HIV. Note that the error term  $\epsilon$  is assumed to be independent and identically distributed with mean 0 and constant variance.

## Discussion

This analysis determined that responses to HIV therapy do vary according to stage of disease, and those with less advanced HIV disease respond better to treatment. In addition, there is a causal relationship between responses to HIV therapy and the extent of prior HIV drug exposure, and there is a positive prolonged clinical benefit from drug intervention in patients with less advanced HIV disease. However, there are discrepancies between the different variables, which raises some concerns for the data collection methods used for this clinical trial. In addition, there are no explanations for why some patients went off their treatment and/or do not have records of their CD4 cell count at 95 weeks, so there may be some bias in the responses to HIV therapy for the patients that continued the clinical trial through 20 weeks and the full 96 weeks. In addition, the censoring effect and the effect of CD8 cells at 96 weeks is assumed to be nonconfounding. However, since there are no variables that track CD8 cells at 96 weeks and no variables that indicate why patients were censored, this study is limited in the generalization it can make. Information about why patients dropped out of treatment would allow for more confident causal conclusions and more clarity about if these assumptions are valid.

## Appendix

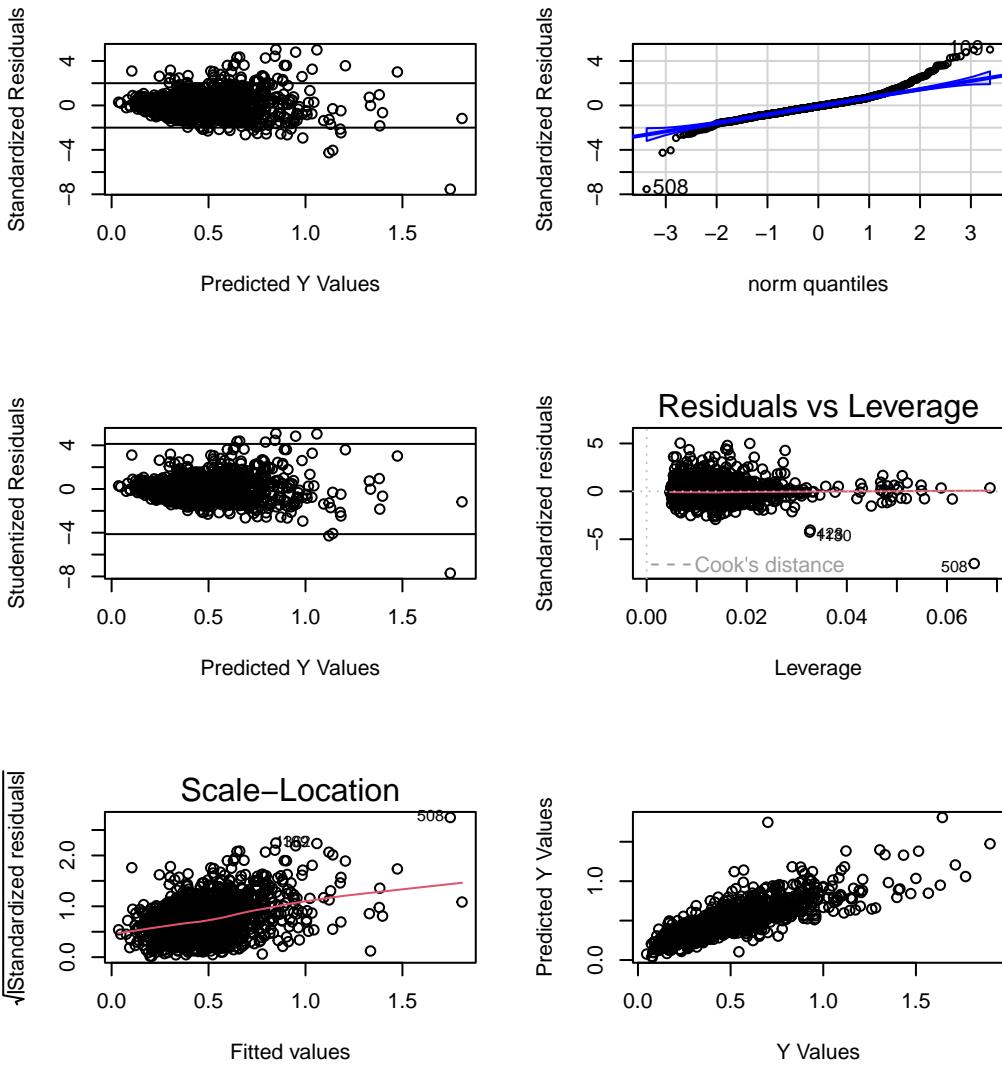


Figure 7: There is no strong correlation between the residuals and the fitted values for the first fitted model. The errors are approximately Normal according to the QQ plot with some deviation in the tails.

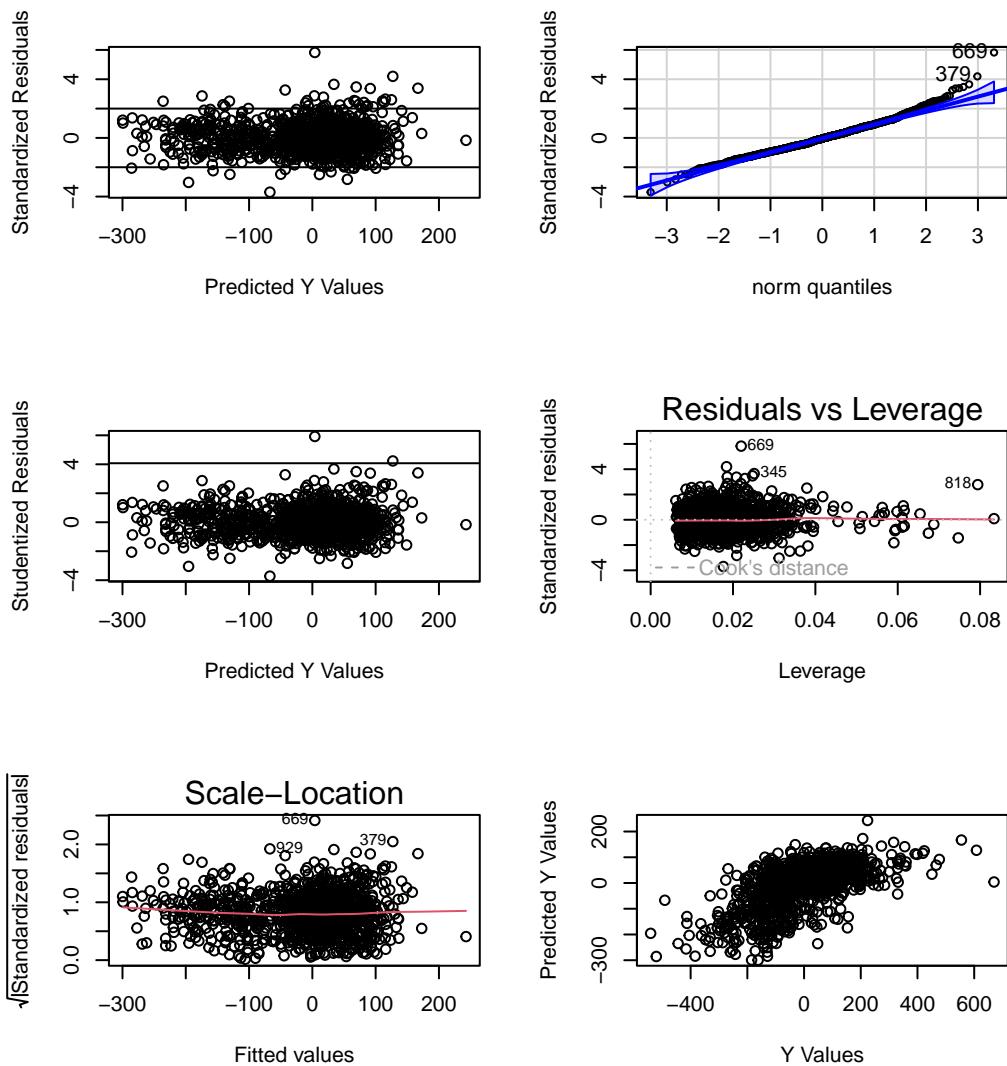


Figure 8: There is no strong correlation between the residuals and the fitted values for the second fitted model. The errors are approximately Normal according to the QQ plot.