Long-term Efficacy of Sub-Silicone Oil Triamcinolone Acetonide on Progression of Proliferative Vitreoretinopathy: A Statistical Study

Md. Anees Parwez **Bhaskar Ray** **Souhardya Sengupta**
Indian Statistical Institute, Kolkata

August 2020

Abstract

*In this article, we carry out a statistical analysis of the efficacy of application of intravitreous triamcinolone acetonide (IVTA) in silicone oil-filled eyes of patients who have undergone relaxing retinectomy as a treatment of proliferative vitreoretinopathy (PVR). We adopt a variable selection methodology for our restricted sample size setting. We also describe some tests for detecting possible 'Selection Bias'. We establish that application of IVTA does result in better outcomes. Also, as a by-product of our analysis, we establish that non-*360° *IRC categories of PVR have similar impact in the long run, if seen as a pre-condition affecting surgery outcomes.*

Keywords: Variable selection, Proliferative Viteroretinopathy, Triamcinolone Acetonide

1 Introduction

Intravitreous triamcinolone acetonide (IVTA) injection [12] is becoming increasingly popular in treating macular diseases. Proliferative vitreoretinopathy (PVR)[3] is a complication that develops in $5 - 12\%$ of patients with rhegmatogenous retinal detachment (RRD)[2] who undergo primary surgery. It hinders the surgical repair of RRD. Relaxing retinectomy (RR)[15] is often done in order to manage such cases. In addition, intravitreal injection of triamcinolone acetonide (TA) in silicone oil-filled eyes can be used to treat cases of PVR. In this study, we are interested in evaluating the long-term efficacy of intravitreous application of triamcinolone acetonide in silicone oil-filled eyes on progression of PVR after 360° RR in patients with chronic retinal detachments.

The analysis has been performed based on BCVA (see Appendix B) values recorded prior to operation, post operation and after follow-up duration as the response variables as well based on whether Silicone Oil Removal (SOR) (Appendix B) was possible or not. The other attributes recorded for each patient were whether IVTA has been applied, their Serial Number, Name, Eye (Left/Right), Indication, PVR, Surgery, Lens Status, Surgery (Primary/Secondary), Cataract Surgery, Complications, Follow-up duration, SOR and NCT.

Our analysis pertains to a low sample size setting, with comparable number of attributes. We develop a variable selection approach to guard against over-fitting and arrive at results. In this article we establish the efficiency of IVTA in PVR surgery, suggest some models for prediction of BCVA scores as well as the odds-ratio of performing SOR. We also make some remark on the severity of the various PVR conditions based on empirical evidences. Rest of the article is organised as follows. In the next section, we describe the dataset, followed by a detailed description of our methodology in Section 3. We conclude in Section 4 and summarise the key observations of our analysis. A quick summary of the medical terminologies can be found in Appendix B.

2 The Dataset

Our data consists of 24 instances of observations made on patients, who underwent a surgery for Proliferative vitreoretinopathy[3], out of which application of the IVTA injection took place in 13 subjects, while the rest 11 were treated with the conventional procedure without IVTA. We have 19 recorded attributes for each of the instances.
In our analysis, we have rejected the following attributes, which we believe do not have any significant impact on the results: Serial Number, Name, Eye (Left/Right), Indication, Surgery, Cataract Surgery and Complications. This reduces our dataset to 11 attributes. A brief description of these attributes follow:

1. **IVTA**: An indicator as to whether there has been application of the IVTA injection. It takes value 1 for 13 instances and 0 for the remaining 11.
2. **PVR**: This variable indicates the condition of the patients pertaining to Proliferative vitreoretinopathy and has four levels: 360° IRC, D2, D3 and Severe Anterior PVR. More information on PVR and it's categorisation can be found in [7]. However, for our purpose, it suffices to note that 360° IRC happens to be the least severe and Severe Anterior PVR happens to be the most severe category out of these four. Table 1 gives the distribution of these levels:
3. **Age**: A continuous variable denoting the age of the subject
4. **Sex**: Gender of the subject. The attribute has two levels: Female and Male having 8 and 16 instances respectively.

Category	Frequency
360° IRC	6
D2	4
D3	8
Severe Anterior PVR	6

Table 1: The distribution of various PVR categories

5. **BCVA Preop**: The BCVA value, prior to the operation of the subject.
6. **BCVA Postop**: The BCVA value immediately after completion of the surgery
7. **BCVA Last**: The BCVA score post-surgery after a followup period. The followup period is different for different subjects, depending on the time taken for stabilisation in the BCVA value for each.
8. **Lens Status**: A categorical variable denoting the Lens status of the individuals. This has four levels: Cataract, Phakic, Pseudophakia, Aphakia. (Please see Appendix B for details on these levels.)
9. **Followup**: A continuous variable denoting the followup period required for each of the subjects, for their BCVA values to get stabilised after the surgery. The period is recorded in months.
10. **SOR**: A categorical variable denoting whether Silicon Oil Removal could be performed or not after the surgery. See Appendix B for explanations.
11. **NCT**: A continuous variable denoting the Intraocular Pressure (IOP) which is measured by the Non-Contact Tonometry (NCT) test.

A decrease in BCVA value is associated with a better outcome. Our objective in this analysis is to decide whether the application of IVTA injection has any significant effect on the outcomes of the surgery. Hence, out of the variables mentioned, we can consider the following two responses:

- Response 1: BCVA Preop - BCVA Last
- Response 2: BCVA Preop - BCVA Postop

It is however a fact that there might be some momentary improvements in the outcome of a surgery immediate to it's completion and the patient's condition might change (progress/deteriorate) in the long run. Hence, Response 2 might be misleading. Hence, we choose to use Response 1 for our analysis, which we shall refer to as '*Response*' in the remaining of this article.

2.1 Challenges

There are however, some preliminary problems that hurdles our analysis. The first being that we have a small sample size compared to the number of attributes. This poses a potential problem that any model we propose for the data might overfit and may not be reliable to draw conclusions for the population.

Another problem that needs consideration is that patients by virtue of their age, the severity of their existing PVR condition or due to any pre-condition might inherently show better results after a surgery. In such cases, even if the effect of IVTA were insignificant, it's effect getting confounded with the inherent better 'recovery capability' of the patients would lead to misleading conclusions. We also have the possibility that probably the IVTA injections were given to patients, who were at a better condition to begin with, a problem referred to as 'Selection Bias'.

Thus, our problems regarding our analysis consist of:

1. Developing a procedure that would avoid the shortcomings of small sample size
2. Testing for the presence of selection bias, and implementing a correction for the same.

3 Statistical Model

3.1 Preliminary Analysis

We begin our analysis by ignoring all other variables and only considering the Response. We want to test whether there has been any significant increase in the response for the patients who were given IVTA injections. Let's make the following assumptions:
Let $Y_1^{(1)}, \ldots, Y_{13}^{(1)}$ denote the responses for the individuals with application of IVTA injection and $Y_1^{(0)}, \ldots, Y_{11}^{(0)}$ denote the responses of those without it. Then we can assume that

$$Y_1^{(0)}, \ldots, Y_{11}^{(0)} \overset{iid}{\sim} f(z),$$
$$Y_1^{(1)}, \ldots, Y_{13}^{(1)} \overset{iid}{\sim} f(z-\theta),$$

where f is continuous and symmetric about 0. To test whether IVTA has an effect, we can test:

$$H_0 : \theta = 0 \text{ Vs. } H_1 : \theta > 0 \tag{1}$$

We can use Wilcoxon Rank Sum test [17] to test the above hypothesis. For this, we have used the function *wilcox.test* available in the R base interface. The following is the outcome:

$$pvalue = 0.02295$$

Thus, such a low pvalue indicates towards the significance of IVTA. This leaves us with the task of identifying any possible 'Selection Bias' and correcting for the same, if any. A linear model seems the best choice for us. However, lack of an appropriate sample size restricts our analysis immensely. For example, it's unlikely that we will be able to incorporate higher order interaction effects. In such a situation, we propose the following: Select variables that best describe the the model to construct the regression line, use a similar methodology to test the significance of IVTA (i.e. choose the best possible variable subset omitting IVTA, and test whether addition of IVTA is significant in the determined best variable subset). The analysis of the variable 'SOR' is done separately. All these are done in the following sections:

3.2 Variable Selection for regression

In this section, we pursue the problem of variable selection for our preliminary regression analysis. Since, our sample size is small, any analysis based on large number of attributes can be potentially misleading. Also, any previous reduction of levels might further help in this endeavor. We begin by making the following observations:

Let's consider the reduction of levels first. Consider the category PVR. Our prior knowledge suggest that '360° IRC' is the least severe, while 'Severe Anterior PVR' is the most severe condition. We introduce two dummy binary variables: 'Dum 1' and 'Dum 2' which are respective indicators of the whether the PVR condition of a patient is non-360° IRC, and whether the PVR condition of the patient is Severe Anterior PVR and we drop the attribute 'PVR'. Note that 'Dum 2' is the additional effect of 'Severe Anterior PVR' over the non-'360°IRC' categories.

Also the four levels of the factor 'Lens Status' is really not that important. We can very well, replace it by a dummy indicator for whether the status of the patient was 'Cataract' or not. Let's call this 'Dum 3' and replace the 'Lens Status' attribute with it.

Now, in order to choose the optimal set of variables, we perform Best subsets regression. Best subsets regression[11], considers all possible models with all possible number of parameters and chooses few optimal ones based on some criteria. For our analysis, we have used the function *regsubsets()* from the R package '*leaps*'[5] and for a visual plot, the function *subsets()* from the package '*car*'[10]. We have used the Bayesian Information Criterion (BIC) as an index for validation for the models. BIC maintains a trade off between good fitted and models with less number of parameters. It's value tend to be lower for those models that explain the best with minimum number of parameters. The plot for the corresponding BICs of each model is given in Figure 1.

As can be seen from the above figure, the model with only two variables: IVTA and Age has the lowest BIC. However, our prior knowledge suggest that the condition factor PVR

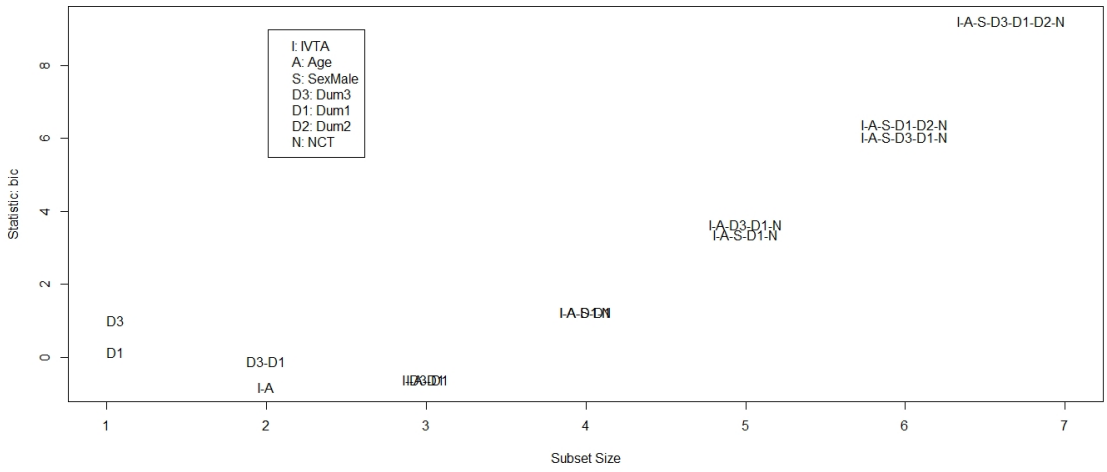


Figure 1: Plot of BIC for the various selected models

(later replaced by Dum 1 and Dum 2) might have a significant impact, as that is the prime cause of carrying out the surgery in the first place. Thus, we would consider the model that contains Dum 1 and/or Dum 2 and has the lowest BIC score. We find two such models with exactly the same BIC:

- **Model 1**: Response $\sim$ IVTA+Age+Dum 1
- **Model 2**: Response $\sim$ IVTA+Dum 3+ Dum 1

3.3 Formulating the Model

Now we observe that for both the selected models, the variable 'Dum 2' doesn't appear. In fact, the variable 'Dum 2' appears with a high BIC value and hence, suggests that selection of this variable is a bad choice. Now, note that 'Dum 2' represented the additional (negative) effect of 'Severe Anterior PVR' over non-'360° IRC' conditions. We also know that 'Severe Anterior PVR' happens to be the most severe PVR condition among the four. Hence, if it's *additional* effect is not significant, then we can very well suggest that there are only two *significant* levels for the factor PVR- presence or absence of the condition '360° IRC.

In order to verify that, we argue as follows: We chose those subsets that had the lowest BIC among all possible subsets, and hence among all those subsets that do not include 'Dum 2'. Now, if we forcefully add 'Dum 2' to these models, and test for it's significance, we can draw conclusion regarding this *additional* effect due to 'Severe Anterior PVR'. We

forcefully add 'Dum 2' to the two models and see the output. (We will use the function *lm()*, available in R base). The pvalues for the significance of the coefficients are given in Table 2. Note that on forcing the variable 'Dum 2' to both the models, the pvalue for

For Model 1

Coefficients	**pvalue**
Intercept	0.3812
IVTA	0.0501
Age	0.1083
Dum 1	0.1819
Dum 2	0.9320

For Model 2

Coefficients	**pvalue**
Intercept	0.00176
IVTA	0.11156
Dum 3	0.10587
Dum 1	0.13609
Dum 2	0.83544

Table 2: Significance of coeff. for adding Dum 2 to Model 1 and 2

testing it's significance turns out to be very high. Thus, we straightaway conclude that the effect of 'Severe Anterior PVR' can be assumed to be similar to the other two categories: 'D2' and 'D3'. Noting that 'Severe Anterior PVR' happens to have the most severe effect, we conclude that all the remaining three levels (other than 360° IRC) can be clubbed into one. Hence, only estimating the effect of $360° IRC$ suffices.
Let's replace 'Dum 1' by a binary variable indicating whether a patient had the '360° IRC condition or not. We will call this variable 'IRC'. We can drop 'Dum 2'. Also, let's rename 'Dum 3' as 'Cataract'.
The summary for the two models with the modified variables are given in Table 3.

For Model 1

Coefficients	**Estimate**	**pvalue**
Intercept	-0.087396	0.7969
IVTA	0.648796	0.0412
Age	0.011787	0.0808
IRC	0.591288	0.1197

For Model 2

Coefficients	**Estimate**	**pvalue**
Intercept	0.6643	0.0238
IVTA	0.5314	0.08411
Cataract	-0.5621	0.0808
IRC	0.07388	0.0739

Table 3: Summary of the two models

We immediately observe that the non-intercept coefficients are all significant at level 10% for Model 2, while none are significant at level 5%. Whereas, for Model 1, the effect of IVTA is significant at 5% level, that of Age at 10%, and that of 360° IRC is insignificant at both the fore-mentioned levels. Hence, from Model 1, the variable 360° IRC can be dropped. The modified summary of the two models are given in Table 4.

As can now, be seen, the effects of both IVTA and Age is significant at level 5% for Model 1. Thus, we choose these coefficients for Model 1. Also observe that from Figure 1, it can be seen that this model has the lowest BIC among all possible ones.

For Model 1

Coefficients	Estimate	pvalue
Intercept	-0.151577	0.6654
IVTA	0.746579	0.0222
Age	0.015763	0.0181

For Model 2

Coefficients	Estimate	pvalue
Intercept	0.6643	0.0238
IVTA	0.5314	0.08411
Cataract	-0.5621	0.0808
IRC	0.07388	0.0739

Table 4: Modified summary of the two models

The models with the renamed/modified variables are:

- **Model 1**: Response $\sim$ IVTA+Age
- **Model 2**: Response $\sim$ IVTA+Cataract+ IRC

The above two models can be used for regression.

3.4 Significance of IVTA

The two linear models formulated in previous section are utilized to make predictions. The motive was to choose the model that fits well. The p-values listed in Table 3 are only an indicator that in those selected models, IVTA is significant (at 10%). However if the motive is to find if IVTA significantly describes the data, we should be looking for model which do not have IVTA, and check if addition of IVTA provides any *significant* improvement.
The implementation of this idea involves dropping the IVTA variable and the finding the model that *best fits the data*. Next, we should augment the predictors in this *best fitting model* by adding the IVTA variable. Finally we need to check if this new IVTA adds *significantly*.
Dropping the IVTA variable, best subset regression yields the best model (lowest BIC) as:

$$response \sim dum1 + dum3$$

. The regression result of adding IVTA to this model is summarized in table:

Variable	Coefficient	pvalue
IVTA	0.5314	0.084
dum1	-0.6618	0.0738
dum3	-0.5621	0.0808

From above we conclude that at 10% significance level, coefficient of IVTA is significant. Also, the BIC associated with the two models are as follows:

Model	BIC
Without IVTA	63.383
With IVTA	62.892

The above table clearly shows a decrease in BIC value with addition of IVTA (of course, this is not the best model with IVTA, as they are already listed in the previous sections). We now observe the fact that IVTA does have a *positive* effect in the change in BCVA scores. In such scenarios, a two-sided t-test for the significance of the coefficients (as we did in this analysis) doesn't take advantage of the full knowledge we have about the data. Thus, if β_{IVTA} denotes the coefficient of IVTA in the above model, the following test should have a higher preference:

$$H_0 : \beta_{IVTA} = 0 \text{ Vs. } H_1 : \beta_{IVTA} > 0$$

The p-value for this test is half of what we reported above. Thus the new pvalue for testing the significance of IVTA is:

$$pvalue = 0.042$$

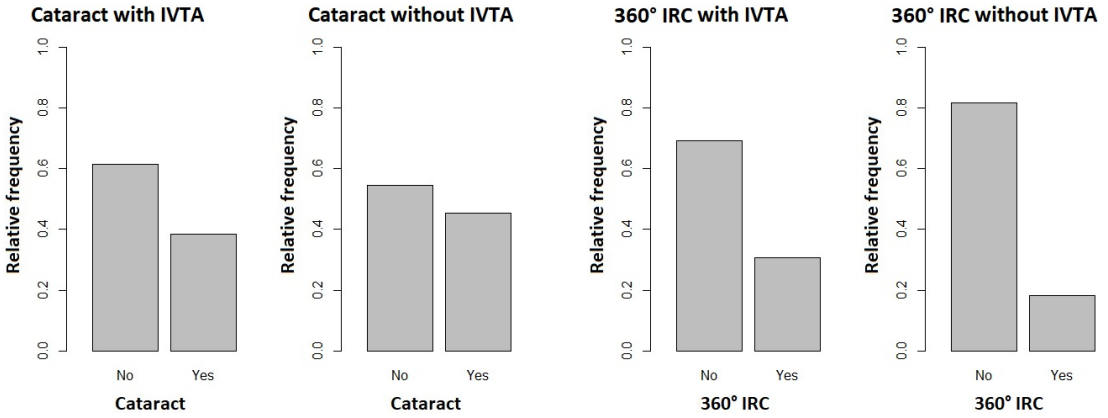
Thus, we conclude that the effect of IVTA in reduction of BCVA scores post surgery, is indeed significant at 5%. Next we check for Selection bias in the various variables of interest we obtained in our above analyses.

3.5 Selection Bias

We described *Selection Bias*, in our context in Section 2.1. We have identified four variables of significance, namely: IVTA, Age, IRC and Cataract. We want to test the significance of IVTA, thus it suffices to check for selection bias in these remaining three variables. Our prior knowledge suggests that we might also look into any possible selection bias for the variable NCT (this will also be evident when we consider the analysis of SOR). Thus the four variables we will check are: Age, IRC, Cataract and NCT.

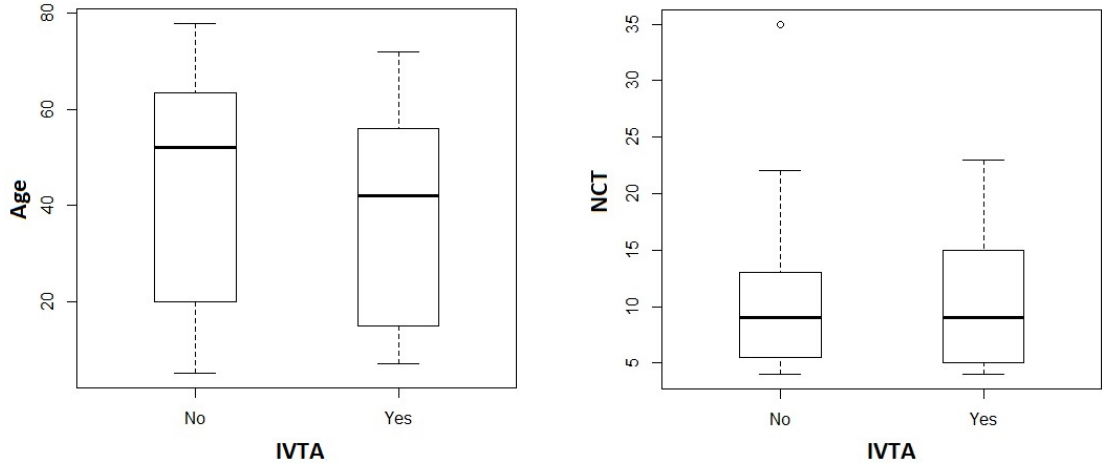
Here we would like to point out that due to small sample size, tests will have very low power. Thus 'non-rejection of null' would be extremely misleading. Also, in such a situation, visual plots should serve the best in drawing conclusions. In this subsection, we provide box plots for the covariates, comparing the distribution for the instances with application of IVTA with those without. For the two binary variables, we show the corresponding bar diagram for the same. Any apparent visual heterogeneity would indicate a possible selection bias.

Figure 2 and Figure 3 contains the bar diagrams and the box diagrams respectively. As can be seen from the figure, there is no apparent heterogeneity for the two categories. Hence, we don't really detect a selection bias visually. We can carry out some exact tests, to



(a) For variable 'Cataract' (b) For variable 'IRC'

Figure 2: Bar plots for Cataract and IRC. For each bar diagram, the left bar represents the frequency in instances without IVTA and the right one with IVTA



(a) For variable 'Age' (b) For variable 'NCT'

Figure 3: Box plots for Age and NCT. For each box plot, the left box represents the instances without IVTA and the right one with IVTA

verify this. However, we should keep in mind that these tests would have very low power due to the small sample size. Rejection will indicate a departure from homogeneity, but acceptance of the null is extremely inconclusive.

For the continuous variables (NCT and Age), we use a permutation test [16] for the hy-

pothesis:

$$H_0 : \text{Distribution is homogeneous for with and without IVTA case} \tag{2}$$

The test is described in Appendix A.
The pvalues of the tests are tabulated below:

Variable	**pvalue**
Age	0.567662
NCT	0.671453

Table 5: P-values of permutation tests

For the categorical variables (Cataract and IRC), for testing the hypothesis(2), we carry out two tests: Fischer's Exact test and Pearson's Chi square test with continuity correction[13]. The later is an approximate test with the added fact that we do not have a large sample, so the large sample approximation might not be that good. The pvalues for the two tests are tabulated below:

Fischer's Exact test

Variable	**pvalue**
Cataract	1.000
IRC	0.649

Chi sq. test

Variable	**pvalue**
Cataract	1.000
IRC	0.813

Table 6: Summary of the tests for categorical variables

We have used the functions *chisq.test()* and *fisher.test()* available in R base, for the above tests. The dataset and all the R codes for reproduction of the results are available at: `https://github.com/bhaskarray99/Effect-of-IVTA`

As the pvalues of the various tests and the plots indicate, we do not seem to have sufficient indications from the data, that the distribution of the variables significantly differ for the with and without IVTA case. Hence, we couldn't detect the presence of Selection Bias w.r.t the variables, we chose. Hence, we can draw conclusions from the results of the linear models we considered, without any adjustment for the bias we suspected.

3.6 Observations and Outcomes

Based on our tests in the previous sections, we developed two models for regression and one for testing the effect of IVTA. Since, we couldn't detect a selection bias, the conclusions we drew from these models are valid directly, without any adjustments. Hence, we conclude

that **IVTA does have a significant positive effect in outcomes of surgery for PVR**.

Further,we propose the following two lines for prediction of the response (difference between BCVA value prior to operation and that after operation, once stabilisation is achieved):

$$Response = -0.152 + 0.746 \times \mathbf{1}_{\{IVTA\}} + 0.016 \times Age. \tag{3}$$

And,

$$Response = 0.664 + 0.531 \times \mathbf{1}_{\{IVTA\}} - 0.562 \times \mathbf{1}_{\{Cataract\}} + 0.074 \times \mathbf{1}_{\{IRC\}}. \tag{4}$$

We however, suggest using (3) as the coefficients show significance at 5% for that equation. In case, the user believes that the PVR condition has an effect on the response and wants a model that incorporates that, (4) can be a choice in such scenarios.

3.7 Analysis of SOR

As we discuss in Appendix B, Silicon Oil Removal (SOR) can be seen as a measure of betterment. The patients for which SOR is a possibility, are likely to be in a better condition. Hence, this demands for a separate consideration of SOR as a response and a thorough analysis for the same. Now, in the dataset available to us, the SOR comes in three levels: 'Done', 'Note Done' and 'Awaiting'.
We observe that it only makes more sense to club the categories 'Done' and 'Awaiting' together into a positive response (which we code with, say '1') as the both indicate the possibility of SOR. The remaining category gets coded as '0'. Thus, to sum up, we obtain a binary response, which we shall call response 3:

- **Response 3:** Whether SOR was possible

This is a binary response as thus we decided to keep it's analysis separate from the previously considered continuous responses. Our approach in this remains the same as before. We first try to find a prediction formula and then test for significance of IVTA.

3.7.1 Variable selection for classification

We adapt a similar approach in analysis Response 3 to what we considered in our previous Response 1. The problem of restricted sample size and large number of attributes still persist and thus a variable selection methodology is the most suitable approach for this case as well.
The variables we consider are exactly the same as we considered in Section 3.2. However for a logistic regression, we consider the *bestglm()* function available with the package *bestglm*[14] available in R. We use the BIC for evaluation of various model. This information criteria is also the default choice of the function. The summary of the best model chosen

Variables	Estimate	Std. Error	z value	P-value
Intercept	-713.02	293536.40	-0.002	0.998
IVTA	615.17	253967.00	0.002	0.998
NCT	19.71	8164.567	0.002	0.998

Table 7: Summary of the best model chosen in logistic regression of Response 3

is as follows:
However, we see that some of the fitted values become very close to 1 (the dataset is available for reproduction of any results) for some of the responses. We also note that there is no issue with the convergence of Fisher Iterations, as the function indicates no such warnings. We thus suspect that probably the data space is linearly separable between the two SOR variants.

3.7.2 Linear separability of the Data Space

The best fitted model we obtained is:

$$\pi(P(\text{Performing SOR})) = -713.02 + 615.17 \times \mathbf{1}_{\{IVTA\}} + 19.71 \times (\text{NCT}), \tag{5}$$

where π is the logit transformation.
In the previous sections, we already tested for presence of selection bias in the variable 'NCT' and found none to be present. Thus, we can use the above model for prediction of the odds-ratio for the ability to perform SOR.
Now consider Figure 4 for a visual depiction of the data space. The Y-axis denotes NCT value for the subjects. The X-axis has two codes : 1 and 0 corresponding to whether IVTA was administered or not. The red dots signify those case where SOR was not possible, and the blue dots those where SOR could be done or is awaiting (which, as we discussed before, are clubbed into a single category). The only green dot is a case that lies in the intersection of both the levels of Response 3. We have one instance bearing the NCT value and IVTA status corresponding to the green dot where SOR was possible, and another bearing the same values, where it wasn't

The black line is a plot of the fitted line of the best model we chose. In particular it's equation is:

$$-713.02 + 615.17 \times (IVTA) + 19.71 \times (NCT) = 0$$

We see that this line passes through the green point that lies in intersection of the two levels of Response 3. Also this line separates the blue and the red points into two sub-regions. We further make the observation that the points corresponding to affirmative SOR removal that lie further from the margin have fitted value close to 1, and the points for which SOR removal was not possible lying further from the margin on the other side correspond to fitted values close to 0 ($\approx 2.22 \times 10^{-16}$).

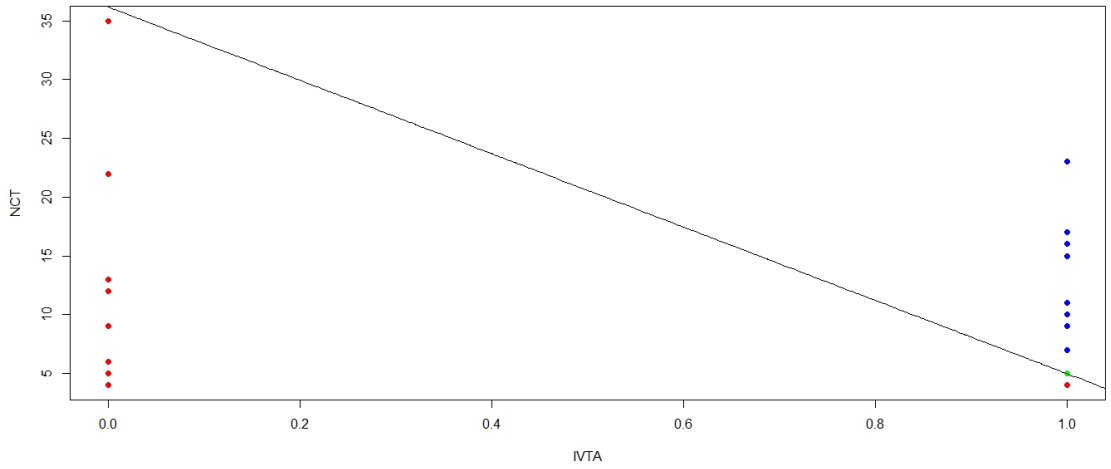


Figure 4: A visualisation of the data space w.r.t to the best chosen variables

3.7.3 Significance of IVTA

The methodology in this section, to test the significance of IVTA is similar to what we considered while testing for Response 1. First we drop IVTA, and then choose the best subset of variables for logistic regression. The best chosen model is:

$$\text{Response } 3 \sim 1(\text{intercept})$$

Now, forcing the IVTA variable into the model, the summary we obtain is given by Table 8.

Coefficient	**Estimate**	**Std. Error**	**P-value**
Intercept	-20.57	5345.91	0.997
IVTA	22.27	5345.91	0.997

Table 8: Summary of the model obtained by adding IVTA to the best subset

Though we see that the pvalues are very high, the reported values of the **null** and **residual deviances** are **33.104** and **11.162** respectively with corresponding **degrees of freedom 23** and **22** respectively. Thus the difference in the deviances are asymptotically distributed as Chi square with degrees of freedom 1, under null assumption. Thus the pvalue for the significance of the above model is:

$$P(\chi_1^2 > 33.104 - 11.162) = P(\chi_1^2 > 21.942) \approx 0$$

Thus this extremely small pvalue indicates that the effect of IVTA is indeed significant for carrying out Silicon Oil removal.

We next consider the apparent contradiction of the very high pvalues of the coefficient, even though the model is significant (with overwhelming confidence). We can produce an explanation for this, considering the high standard error of the estimates. It might be the case that the information number of the coefficients (which is the inverse of the asymptotic variance of the corresponding estimate) is quite low, which results in this high standard error and consequently such high pvalues.
There might be more than one possible estimate of the coefficients that produce the same classification. Hence, we can't really say that this particular estimate is significant (as there might be many more that produce the same results). This is reason we get such a large pvalue. However, as the test of deviances suggest, the effect of IVTA is indeed significant, regardless of which estimate we decide to work with.

3.7.4 Deductions

Thus to sum up our discussions, we first obtained a prediction line for prediction of odds ratio of the ability to do SOR, given by:

$$\pi(P(\text{Performing SOR})) = -713.02 + 615.17 \times \mathbf{1}_{\{IVTA\}} + 19.71 \times (\text{NCT}),$$

where π is the logit transformation.

We also demonstrated the linear separability of data-space through Figure 4. and carried out a test for significance of IVTA and concluded that **IVTA has a positive significant effect in aiding the possibility to perform SOR**.

4 Discussion and Conclusions

In this article, we explored the effect of sub-silicone oil Triamcinolone on progression of Proliferative Viteroretinopathy in terms of both decrease in BCVA values post surgery and ability to perform SOR on the patient. Our analysis was based on a dataset with small number of instances and a considerably large set of attributes. We did proper identification of responses and did variable selection to obtain a subset of the variables to do statistical analysis. This selection was based on a prior elimination based on our previous knowledge followed by a further reduction using best subsets regression. We identified two models that fits the best for decrease in BCVA values as a response. Also we fitted a logistic regression to predict the odds ratio of the ability to perform SOR. We described how selection bias could affect our analysis and also tried to detect possible selection bias among the variable. We used visual and other exact and approximate statistical methods. We couldn't detect

any significant selection bias.

Our analysis shows that sub-silicone oil Triamcinolone does have a significant positive effect on the progression of Proliferative Viteroretinopathy in terms of both the measures we discussed in the previous stanza. Patients who were given this injection, performed significantly better than those who were not. Also as a by-product of our analysis, we drew a conclusion that the most severe PVR category (Severe Anterior PVR) doesn't give significantly bad results as compared to the other non-360° IRC(the least severe) categories, i.e the effects of 'D2','D3' and 'Severe Anterior PVR' are all similar.

References

[1] Measurement of visual acuity. `http://www.ssc.education.ed.ac.uk/courses/vi&multi/vmay06a.html`.

[2] Retinal detachment: Causes & how to get treatment. `https://www.nvisioncenters.com/conditions/retinal-detachment`.

[3] Proliferative vitreoretinopathy, Jan 2020. `https://eyewiki.aao.org/Proliferative_Vitreoretinopathy`.

[4] Francesco Barca, Tomaso Caporossi, and Stanislao Rizzo. Silicone oil: Different physical proprieties and clinical applications, Jun 2014.

[5] Thomas Lumley based on Fortran code by Alan Miller. *leaps: Regression Subset Selection*, 2020. R package version 3.1.

[6] Kierstan Boyd. What is vitrectomy?, Apr 2020.

[7] Salvatore Di Lauro, Mustafa Kadhim, David Charteris, and J Carlos Pastor. Classifications for proliferative vitreoretinopathy (pvr): An analysis of their use in publications over the last 15 years. *Journal of Ophthalmology*, 2016:1–6, 06 2016.

[8] ejb117 and ejb117. Retinectomy, Apr 2017.

[9] Christiane I Falkner, Susanne Binder, and Andreas Kruger. Outcome after silicone oil removal, Nov 2001.

[10] John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition, 2019.

[11] R. R. Hocking and R. N. Leslie. Selection of the best subset in regression analysis. *Technometrics*, 9(4):531–540, 1967.

[12] Jost B Jonas. Effects of triamcinolone acetonide injections with and without preservative, Sep 2007.

[13] Hae-Young Kim. Statistical notes for clinical researchers: Chi-squared test and fisher's exact test. *Restorative Dentistry & Endodontics*, 42:152–155, 05 2017.

[14] A.I. McLeod, Changjiang Xu, and Yuanhao Lai. *bestglm: Best Subset GLM and Regression Utilities*, 2020. R package version 0.37.3.

[15] Khaled Ag Shalaby. Relaxing retinotomies and retinectomies in the management of retinal detachment with severe proliferative vitreoretinopathy (pvr), Oct 2010.

[16] Larissa Stanberry. *Permutation Test*, pages 1678–1678. Springer New York, New York, NY, 2013.

[17] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80, 1945.

A Description of the permutation test in Section 3.5

Here, we give a description of the permutation test mentioned is Section 3.5.
Permutations tests are a class of resampling tests, that considers all possible permutations of the sample to construct a distribution for various testing problems about the population. Permutation tests are a popular choice for small sample sizes when we have no other options to proceed with. Not only that there are better alternatives available for large sample sizes, but permutation tests, being a computationally expensive procedure, is infeasible under such a setting. However, for our low sample size situation, it is one of the best tools we have.
Another thing to be mentioned is that, these tests have low power, especially when we have small sample size. Thus, the probability of committing a type II error is high. However, these test does guard against type I error, like any test, and rejection of the null hypothesis is a sufficiently strong evidence for us.

In our case, we have a total of 24 responses $(Y_1, Y_2, \ldots, Y_{24})$ out of which 13 responses $(Y_1, Y_2, \ldots, Y_{13})$ correspond to the group with application of IVTA. We shall treat the set of 24 responses as the population. Our sample will consist of 13 responses. The pseudocode for the permutation test (including calibration of the distribution and calculation of the pvalue) is given in Algorithm 1.

Algorithm 1: Permutation Test

Set $\Delta = \{\text{i} \mid \underset{\sim}{y}^{(i)}$ is an SRSWOR of size 13 from the 24 available responses$\}$
for *i in* Δ **do**
  Let $\underset{\sim}{y}^{(i)} = \{y_1^{(i)}, y_2^{(i)}, \ldots, y_{13}^{(i)}\}$
  Set $\bar{y}^{(i)} := \frac{y_1^{(i)} + y_2^{(i)} + \ldots + y_{13}^{(i)}}{13}$
  Then,
  $\mathbb{E}\left(\bar{y}^{(i)}\right) = \frac{Y_1 + Y_2 + \ldots + Y_{24}}{24} = \bar{Y}$
  $V(\bar{y}^{(i)}) = \left(\frac{1}{13} - \frac{1}{24}\right) s_Y^2$
  Define, $T^{(i)} = \left(\frac{\bar{y}^{(i)} - \bar{Y}}{\sqrt{\left(\frac{1}{13} - \frac{1}{24}\right) s_Y^2}} \right)^2$
end
Set $T = \left(\frac{\frac{Y_1 + Y_2 + \ldots + Y_{13}}{13} - \bar{Y}}{\sqrt{\left(\frac{1}{13} - \frac{1}{24}\right) s_Y^2}} \right)^2$
p -value $= \sum_{i \in \Delta} \frac{\mathbb{1}(T^{(i)} \geq T)}{\binom{24}{13}}$

B Definitions and Descriptions

Retinal Detachment(RD): This is a medical emergency where the retina of the eye is ruptured and leaves contact with the nourishing blood vessels. This is generally caused by injury or age and might lead to blindness if medical assistance is not provided. There are three types of it:

- Exudative
- Rhegmatogenous
- Tractional

Rhegmatogenous Retinal Detachment (RRD): This is most common form of retinal detachment. The retina breaks and allows fluids to pass beneath [2].

Proliferative Vitreoretinopathy (PVR): (*formerly named "massive vitreous retraction"*).This is a diseased condition formed as a complication of RRD. PVR is estimated in 5-10 percent of RRD. In this proliferative membrane form in vitreous on either side of retina [3, 7].

Vitrectomy: The space between retina and lens is filled by vitreous humor gel. Vitrectomy is a procedure where this gel is temporarily removed to better access the retina. This needs to be performed in severe cases of Retinal Detachment [6].

Retinectomy: This refers to surgical removal of a part of the retina. The central retina is untouched. (This surgery is undertaken in cases where retina would not lie flat due to severe detachment) [8].

Relaxing retinectomy(RR): This is performed in cases of severe retinal detachment, especially when retina is shortened (either due to retinal incarceration or PVR). In this procedure, peripheral retina is excised to preserve function of posterior retina, which is more significant for vision [15].

Silicone Oil (as intraocular tamponade): Silicon oils are used in vitrectomy surgery. They can replace aqueous homour on retinal surface. These are adhesive too and can hold retina fixed to retinal pigment epithelium. This is injected during vitrectomy surgery.
But the long term presence of this fluid in eyes, cause poor vision, complications and is undesirable. Thus removal of silicon oil is recommended for stability [4, 9].

Intravitreal: Refers to administration of drugs into the vitreous humor of the eye. Eg: intravitereal injections.

Intravitreal Triamcinolone Acetonide (TA) crystal: The use of TA crystals for

treatment of PVR was first suggested by Robert Macheme, who is also credited to surgically interact with vitreous humor. This is applied at the time of silicon oil injection in the vitreous cavity, which acts as a reservoir for this drug. It is also used in diabetic Retinopathy [12].

Visual Acuity: Visual Acuity refers how precisely an examinee's eye perceive small details. Optometrists employ Snellen's Chart for Visual Acuity testing.An examinee score of 6/x means that what this examinee can read from 6 meters can be read by a normal eye from x meters(usually $x \geq 6$).

Best Corrected Visual Acuity(BCVA) BCVA represents best possible vision that can be achieved by using corrective measures, such as glasses or lens. In our data, the BCVA is measured as logMAR score, which overcomes the shortcoming of non-uniformity of legibility of different alphabets.

Logarithm of minimum angle of resolution(LogMAR): It measures the visual acuity loss. Thus lower the LogMAR the better the vision. Each letter has score of 0.02 log units , with five letters on each line. Thus each line represents change of 0.1 log unit. A logMAR score of 0.0 represents a healthy 6/6 vision. [1].

Phakic lenses: Literally, it means 'having a natural crystalline lens'. The natural eye lens is not removed and an artificial lens is surgically implanted.

Aphakic lens: This refers to the absence of natural eye lens. Since the natural lens has ability to 'accommodate' depending on object distance, absence of natural lens leads to loss in this power of accommodation. Lens may need to be surgically removed in situations such as cataract.

Pseudophakia: Literally, it means 'fake lens'. It refers to surgical insertion of artificial lens in place of natural eye lens. A cataract could lead to aphakic eye condition followed by insertion of pseudophekiac Intraocular Lens(IOL).

Cataract: A diseased condition of eye, with around 1 million annual cases in India. In this the natural eye lens becomes cloudy as a result of proteins in the eyes breaking down and clumping together. A surgical treatment involves removing the natural eye lens and inserting IOL.