Statistical Analysis for Functional Health Literacy on Treatment Adherence among Type 2 Diabetic Clients' Survey Data

Alfeo Sabay

2020-01-14

Data Description

Survey data was gathered using two tools from 276 Diabetic Clients. The NVS (Newest Vital Sign) tool produced a literacy score ranging from zero to six (0 with the lowest literacy to 6 with the highest literacy) from a series of survey questions presented to the Diabetic Client. Similarly, the BMQ (Brief Medical Questionaire) tool derived a score in the range of zero to eleven (0 as most adherent and 11 as least adherent) relfecting the patient's adherence to treatment from a questionaire.

This survey includes Demographic data. Age, Other Medications and Gender. Age is the disclosed age of the subject. Other Medications (coded as yes/no) reflect whether or not the subject is currently taking other medications apart from the diabetic medication. Gender (coded as Male, Female) reflect the gender of the subject.

In order to examine the possible effects of variables that impact adherence, Logistic Regression will be performed on the data set. As such, the reponse variable BMQ has been coded as "Adherent" for scores of zero and "Non-Adherent" for scores greater than zero. Similarly, a new independent variable (ordinal categorical) was derived from the NVS Scores to reflect 3 literacy levels ("limited", "possibly", "adequate"). NVS scores of 0-1 is coded as "limited" literacy level. NVS scores of 2-3 are coded as "possibly" literacy level. NVS scores of 4-6 are coded as "adequate" literacy level.

Table 1: Table continues below

NVS	Regimen	Belief	Recall	Access	BMQ	Gender	Age
2	1	0	0	0	Non-Adherent	Male	45
6	0	0	1	1	Non-Adherent	Female	47
4	0	0	1	1	Non-Adherent	Male	47
1	0	0	0	0	Adherent	Male	78
0	0	0	0	0	Adherent	Female	67
1	0	0	0	0	Adherent	Female	56

Other_Medications	Literacy_level
yes	possibly
no	adequate
yes	adequate
yes	limited
yes	limited
yes	limited

Table 1: Data Sample from Dataset

Table 3: Table continues below

NVS	Regimen	Belief	Recall
Min. :0.000	Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:2.000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :3.000	Median :0.0000	Median :0.0000	Median :0.0000
Mean :3.315	Mean :0.4239	Mean :0.1413	Mean :0.4891
3rd Qu.:5.000	3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:1.0000
Max. :6.000	Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 4: Table continues below

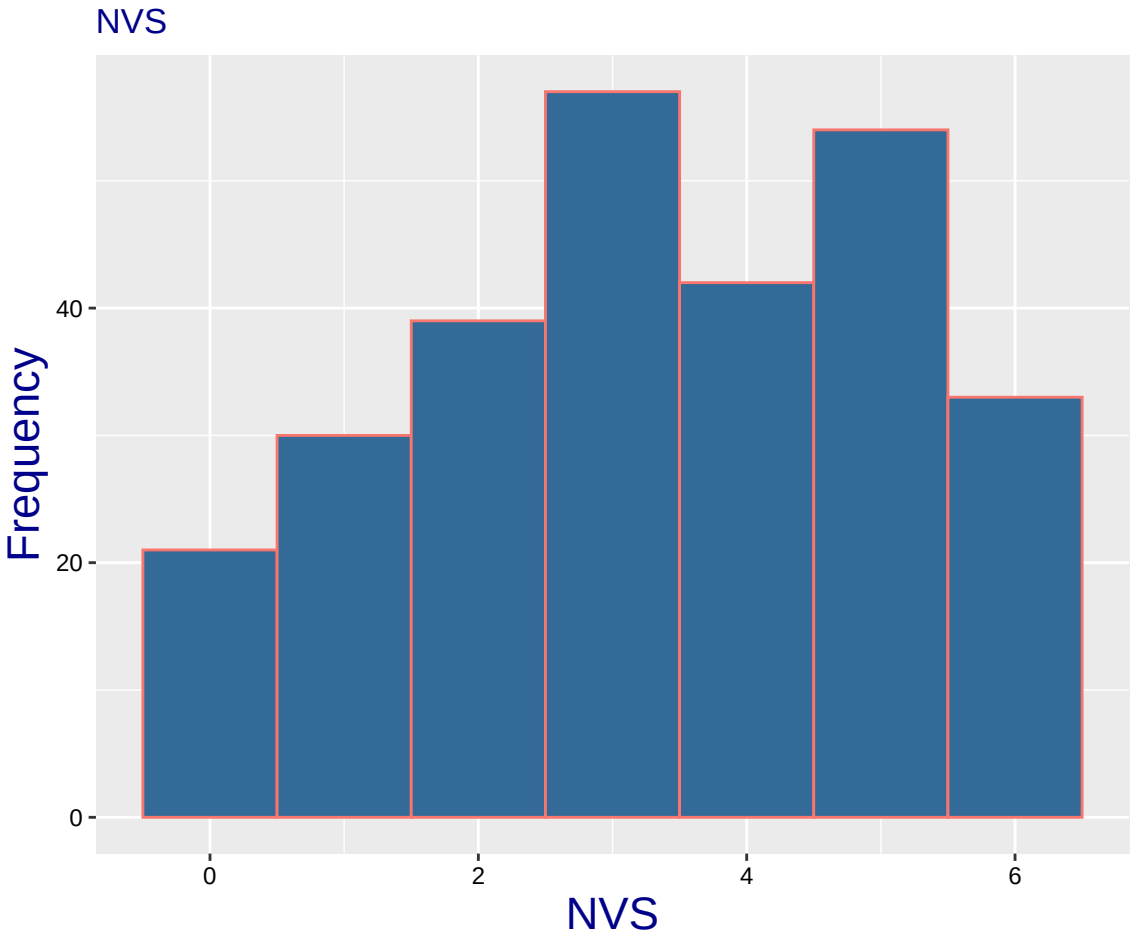
Access	BMQ	Gender	Age
Min. :0.0000	Adherent : 84	Female:165	Min. :37.00
1st Qu.:0.0000	Non-Adherent:192	Male :111	1st Qu.:51.00
Median :0.0000	NA	NA	Median :58.00
Mean :0.4239	NA	NA	Mean :58.55
3rd Qu.:1.0000	NA	NA	3rd Qu.:67.00
Max. :1.0000	NA	NA	Max. :82.00

Other_Medications	Literacy_level
no : 77	limited : 51
yes:199	possibly: 96
NA	adequate:129
NA	NA
NA	NA
NA	NA

Table 2: Statistical Summary

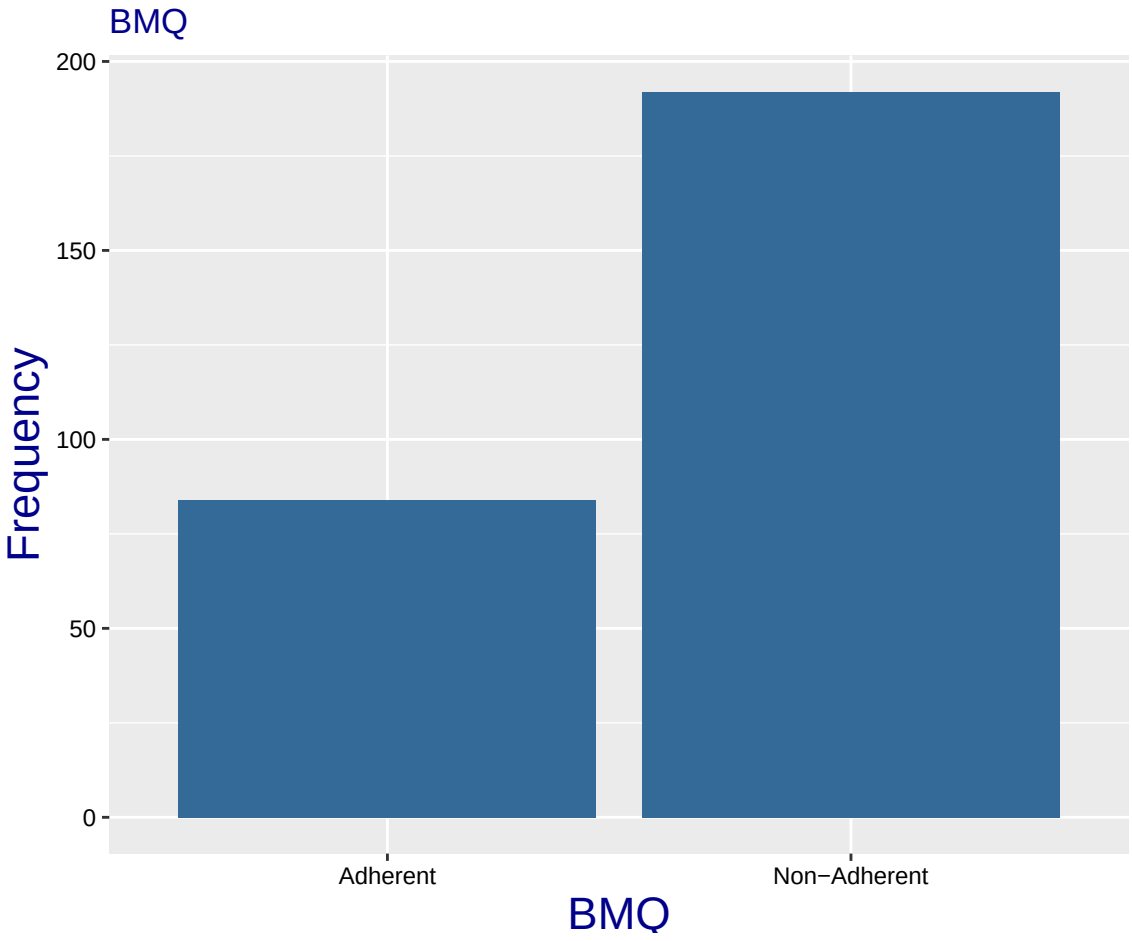
Descriptive Statistics

Plots are shown below for initial exploration of the data.



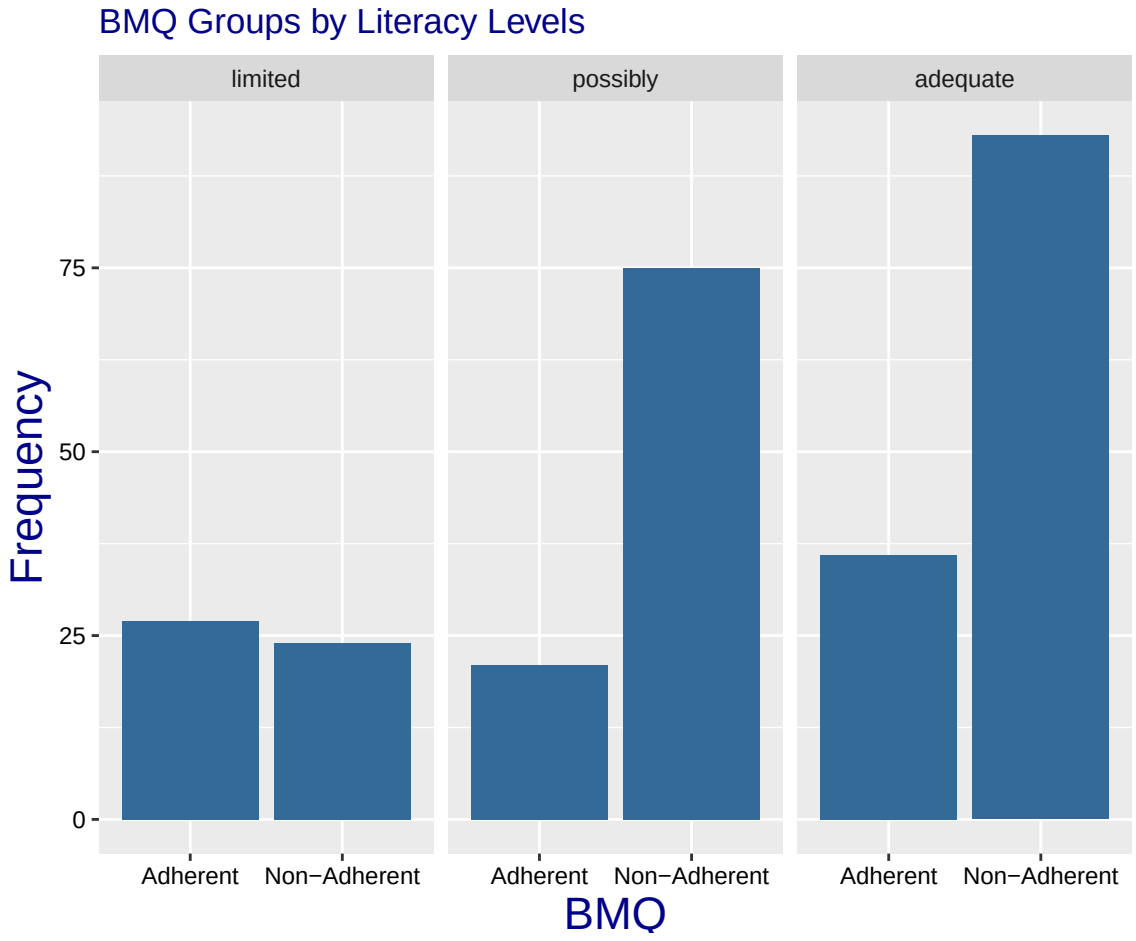
NVS	n
0	21
1	30
2	39
3	57
4	42
5	54
6	33

Figure 1: NVS - Literacy Distribution



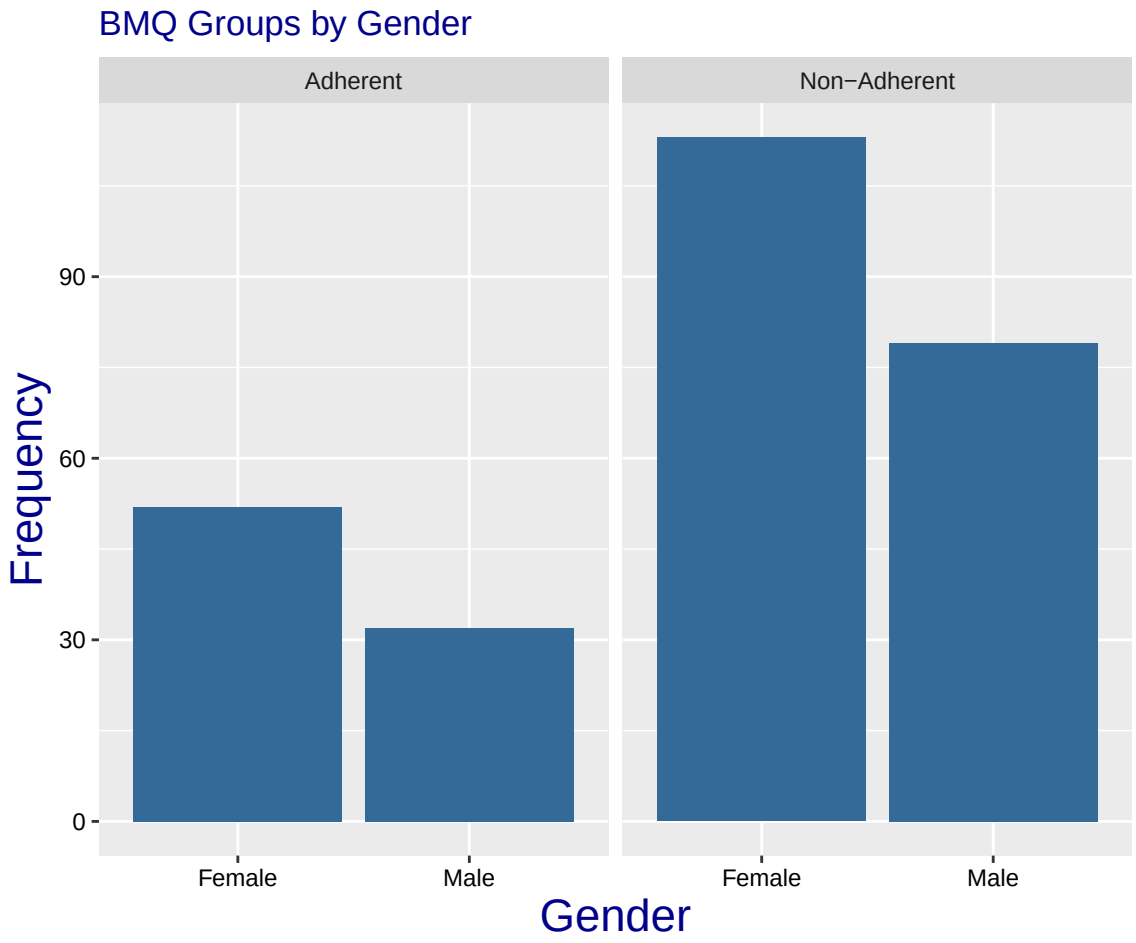
BMQ	n
Adherent	84
Non-Adherent	192

Figure 2: BMQ - Overall Distribution



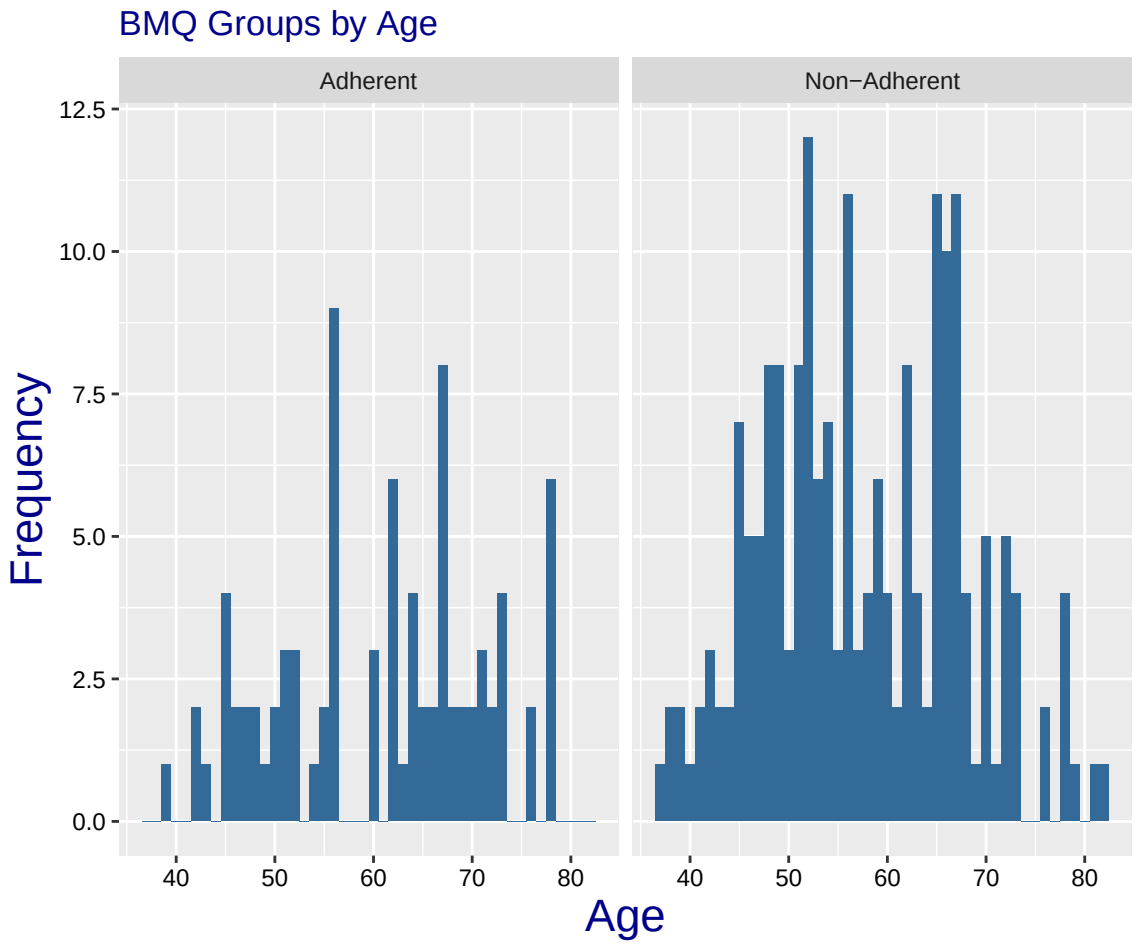
Literacy_level	BMQ	n
limited	Adherent	27
limited	Non-Adherent	24
possibly	Adherent	21
possibly	Non-Adherent	75
adequate	Adherent	36
adequate	Non-Adherent	93

Figure 3: BMQ - By Literacy Groups



BMQ	Gender	n
Adherent	Female	52
Adherent	Male	32
Non-Adherent	Female	113
Non-Adherent	Male	79

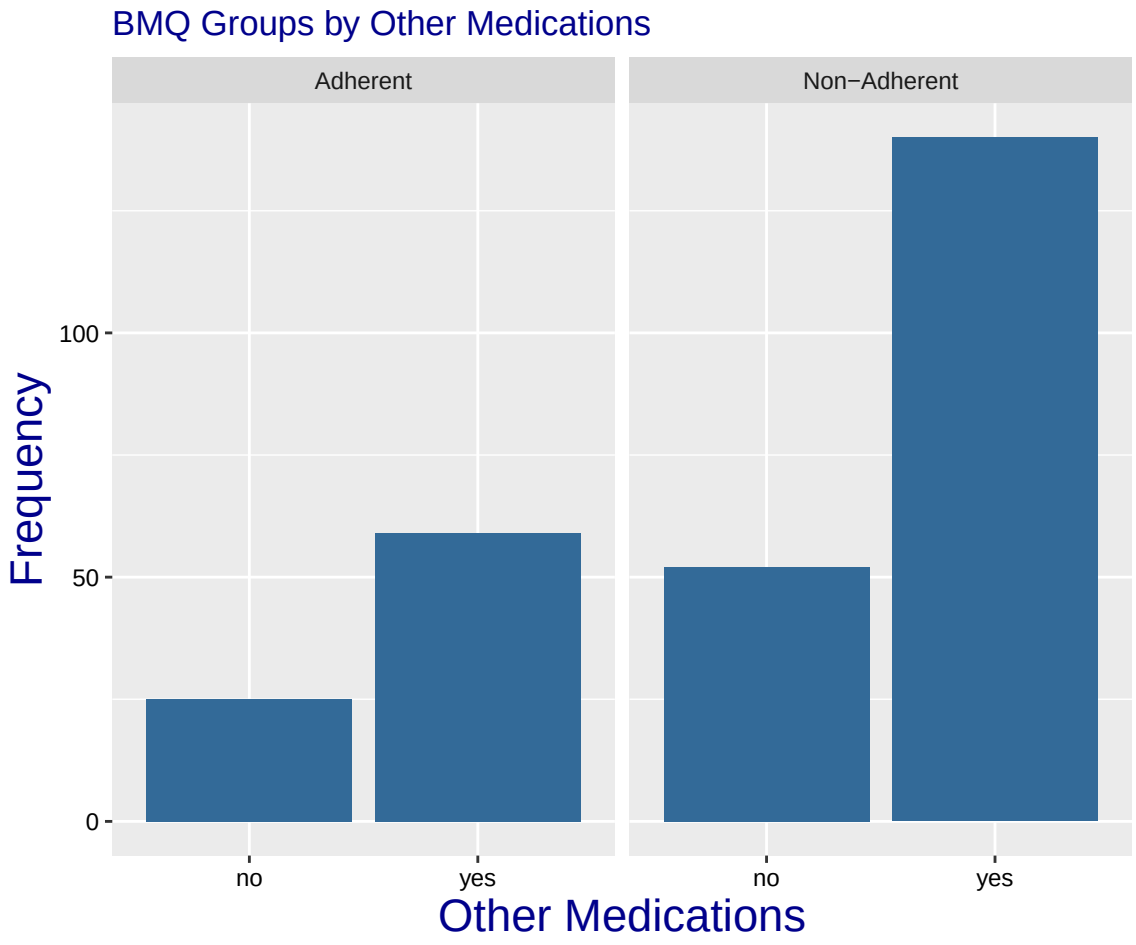
Figure 4: BMQ - By Gender



BMQ	Age
Adherent :84	Min. :39.00
Non-Adherent: 0	1st Qu.:52.00
NA	Median :62.00
NA	Mean :60.79
NA	3rd Qu.:68.25
NA	Max. :78.00

BMQ	Age
Adherent : 0	Min. :37.00
Non-Adherent:192	1st Qu.:49.75
NA	Median :56.00
NA	Mean :57.58
NA	3rd Qu.:66.00
NA	Max. :82.00

Figure 5: BMQ - By Age



BMQ	Other_Medications	n
Adherent	no	25
Adherent	yes	59
Non-Adherent	no	52
Non-Adherent	yes	140

Figure 6: BMQ - By Other Medications

Chi-square Test For Independence (Null Hypothesis: Adherence and Literacy are Independent) The small p-value (0.0003503) in the Chi-square test output below means we reject the null hypothesis that BMQ and Literacy_level are independent of each other and we accept the alternative hypothesis that there is a relation between BMQ and Literacy_level.

Test statistic	df	P value
15.91	2	0.0003503 * * *

Table: Pearson's Chi-squared test: `v_table`

Table 3: Chi-Square Test for Adherence and Literacy

Logistic Regression

We first start with a full Logistic Regression Model.

BMQ = Literacy_level + Gender + Age + Other_Medications

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.611	0.849	3.075	0.002103
Literacy_level.L	0.7967	0.2474	3.221	0.001278
Literacy_level.Q	-0.7242	0.2497	-2.901	0.003723
GenderMale	0.1287	0.2804	0.4591	0.6461
Age	-0.03441	0.01343	-2.563	0.01037
Other_Medicationsyes	0.0962	0.3013	0.3193	0.7495

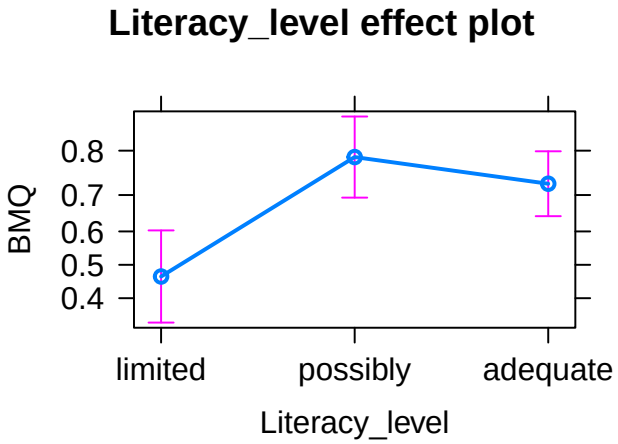
Table: Fitting generalized (binomial/logit) linear model: BMQ ~ Literacy_level + Gender + Age + Other_Medications

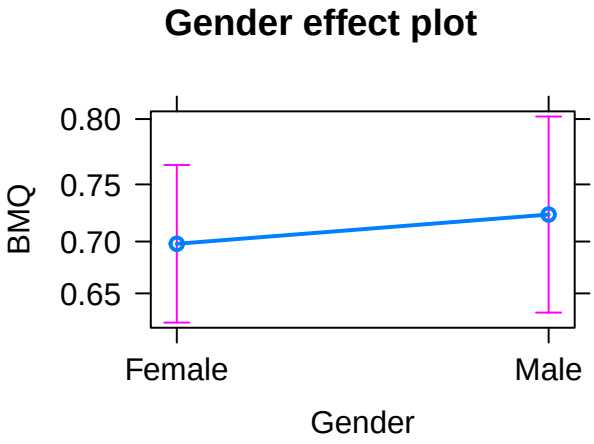
Table 4: Full Model Logistic Regression

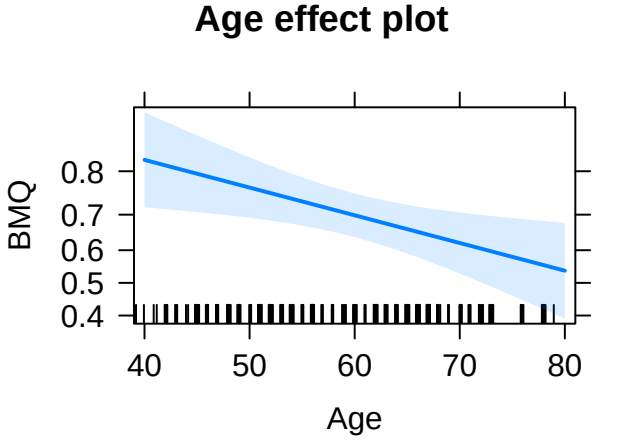
As can be seen from the results of the Logistic Regression above, the p-values for the variables Other_medications (0.7495) and Gender (0.6461) are large meaning these two variables do not contribute significantly to the response variable BMQ.

As such, we can build a simpler model without Gender and Other_medications.

BMQ = Literacy_level + Age







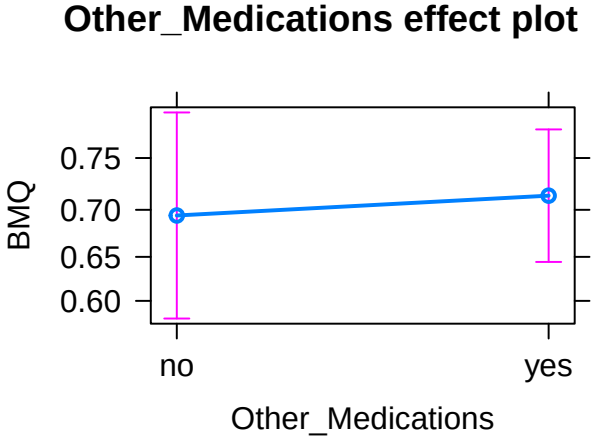


Figure 7: Full Model Effects Plot

As can be seen, the p-values in the reduced model are below the significance level of .05 therefore Literacy_level and Age can be chosen as the predictors.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.73	0.8113	3.365	0.0007646
Literacy_level.L	0.7918	0.2469	3.207	0.001342
Literacy_level.Q	-0.7348	0.2488	-2.953	0.003149
Age	-0.03439	0.01342	-2.563	0.01039

Table: Fitting generalized (binomial/logit) linear model: BMQ ~ Literacy_level + Age

Table 5: Reduced Model Logistic Regression

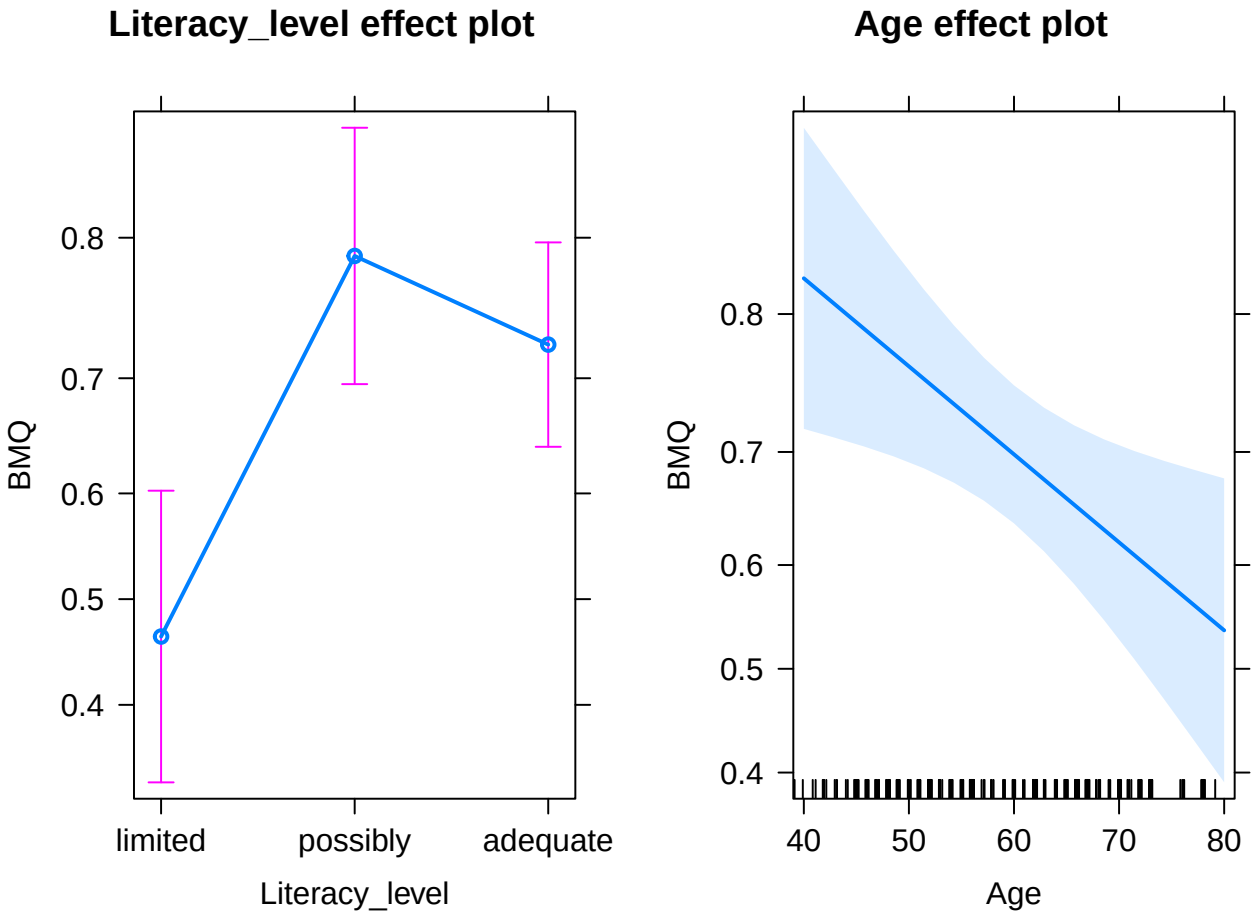


Figure 8: Reduced Model Effects Plot

Since the reduced model is a subset of the full model, we can perform ANOVA (Analysis of Variance) on the two models to see if they perform the same.

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
272	317.3959	NA	NA	NA
270	317.0817	2	0.3141375	0.8546453

Table 6: ANOVA: Full and Reduced Model Comparison

With a p-value of 0.85, we can conclude that the reduced model performs just as well as the full model therefore we will use the reduced model going forward.

Model Coefficients

The coefficients of the fitted model gives us information on the predictor's effect on the response variable BMQ. The values below have been converted to Odds Scale.

The table 7 below shows the coefficients for the reduced BMQ model. Literacy_level.L is the linear coefficient while Literacy_level.Q is the quadratic coefficient. For our purposes, the value of interest is Literacy_level.L. Here it simply means that for every increment of Literacy_level, the effect of BMQ is multiplied 2.207 times and .966 times every increment in Age.

Table 8 shows these model coefficients with confidence interval of 95%.

(Intercept)	Literacy_level.L	Literacy_level.Q	Age
15.34	2.207	0.4796	0.9662

Table 7: Model Coefficients (Odds Scale)

	2.5 %	97.5 %
(Intercept)	3.214	78.04
Literacy_level.L	1.365	3.605
Literacy_level.Q	0.2904	0.7728
Age	0.9407	0.9917

Table 8: Model Coefficients with 95% Confidence Intervals

Examining The Effects of Literacy_level and Age (Table 9) With the fitted BMQ model, we can examine the effect (probability) of each predictor variable by performing a model predict using test data. In this test, we plug in 3 levels of literacy ("limited", "possibly", "adequate") while holding age constant (mean value of Age).

With this test data, it shows that the probability of "limited" literacy is below .5 or 50% which would classify this subject as Non-Adherent while "possibly" and "adequate" levels would classify as Adherent.

Literacy_level	Age	prob	Pred
limited	58.55	0.4642	Non-Adherent
possibly	58.55	0.7886	Adherent
adequate	58.55	0.7264	Adherent

Table 9: Fitted Model Prediction Age Constant

Examining The Effects of Literacy_level and Age (Table 10) In the test prediction below, we created a data set holding Literacy_level at "limited" with Age values of 50, 60 and 70. From Table 10 below, we can see that only the 50 year old subject classifies as Adherent while the 60 and 70 year old subjects classify as Non-adherent (below .5 probability). This is consistent with the Effect Plots above (Figure 8).

Literacy_level	Age	prob	Pred
limited	40	0.6212	Adherent
limited	50	0.5376	Adherent
limited	60	0.4519	Non-Adherent
limited	70	0.3689	Non-Adherent

Literacy_level	Age	prob	Pred
limited	80	0.293	Non-Adherent

Table 10: Fitted Model Prediction Holding Literacy Level At Limited

Examining The Effects of Literacy_level and Age (Table 11) In the test prediction below, we created a data set holding Literacy_level at "possibly" with Age values of 50, 60 and 70. From Table 11 below, all 3 ages classify as "Adherent". This is consistent with the Effect Plots above (Figure 8).

Literacy_level	Age	prob	Pred
possibly	40	0.8759	Adherent
possibly	50	0.8335	Adherent
possibly	60	0.7802	Adherent
possibly	70	0.7156	Adherent
possibly	80	0.6408	Adherent

Table 11: Fitted Model Prediction Holding Literacy Level At Possibly

Examining The Effects of Literacy_level and Age (Table 12) In the test prediction below, we created a data set holding Literacy_level at "adequate" with Age values of 50, 60 and 70. From Table 12 below, all 3 ages classify as "Adherent". This is consistent with the Effect Plots above (Figure 8).

Literacy_level	Age	Prob	Pred
adequate	40	0.834	Adherent
adequate	50	0.7808	Adherent
adequate	60	0.7164	Adherent
adequate	70	0.6417	Adherent
adequate	80	0.5594	Adherent

Table 12: Fitted Model Prediction Holding Literacy Level At Adequate

Post Hoc Power Analysis

Based on a similar study by Olaolournpo O, et. al. 2018, at the start of the study, the sample size was calculated using a Test For Proportions Power Analysis. This resulted in a minimum sample size of 242 subjects given two proportions of .33 (Adequate Literacy, Non-Adherent) and .55 (Limited Literacy, Adherence) with desired Power of .8 at a Significance Level of .05. Based on this sample calculation, the team gathered survey data for 276 subjects.

Later in this study a deeper statistical analysis using Logistic Regression was necessary in order to get a better understanding of the predictor effects on the response variable (BMQ) which represents the patient's adherence. The minimum sample size calculation for Logistic Regression (General Linear Model) differs from that of Test for Proportions as previously calculated thus we recalculated the minimum sample size in order to insure proper sample representation for the Power (.8) desired at a Significance Level of .05.

General Linear Model Sample Size Calculation

f2 = (R1 - R2)/1-R1 - the effect size

f2 = (.55-.33)/(1-.55) = .48

u = Number of Predictors (4)

sig.level = .05

power = .8

Plugging in these values to the pwr.f2.test function (R), results in the following: pwr.f2.test(u=4, f2=.48, sig.level = .05, power=.8)

- **u**: *4*
- **v**: *25.01*
- **f2**: *0.48*
- **sig.level**: *0.05*
- **power**: *0.8*
- **method**: Multiple regression power calculation

The output v is the denominator degrees of freedom. The formula for v is: N - u - 1 = v Calculating for N (sample size for Logistic Regression Model) N = 25 + 4 + 1 = 30 This is the minimum sample size required for the above Logistic Regression. As such the sample size of 276 from the dataset is more than adequate and therefore can only increase the Power of the Prediction Probabilities in this regression.

The plot below demonstrates how Power varies with respect to sample size. The power required in this study is .8 with a calculated minimum sample size of 30.