

# **draft report**

*Chuqiao Yan*

*15/03/2020*

## **Abstract**

This report analyzed the data collected from the Best Rewarded Innovation Inc's health survey through its BLR platform, in order to give a better support for survey respondents whose engaging behaviors relates to a better health outcome. This analysis was designed to investigate the relationship of a health improvement rate and contributable characteristics; also, the relationship of a health improvement and usage frequency of the BLR platform. The analysis indicated that heart score, respiratory score, mental health score, diet score, physical activity score, financial health score, medication score, sleep score, stress score and smoking score have influence on the BLR improvement rate. However, there was not sufficient evidence to show that higher usage frequency will lead to the BLR improve.

## **1. Introduction**

This report is responding to the analysis of data collected from the Best Rewarded Innovation Inc's health survey through its BLR platform. With the goal of better supporting survey respondents whose engaging behaviors in this survey relating to a better health outcome, all survey respondents were collected over three companies: company M; company R and company Z, over 409,767 users engaged during the year of 2015-2019 in total. In general, this BLR platform worked by tracking users' activities within different time period by giving a BLR score output, which was measured by other 17 health characteristics. The first research question was going to find what characteristics of the users was associated with a higher rate of improvement for the health risk factors. Meanwhile, because the BLR platform does have a motivation mechanism when these user logins each time to either gain or lose points, the second research question will address if a higher frequency of usage of the platform associated with a general improvement in health.

The summary of data for manipulation is given in the section 2. The methods for model fitting and results explanation will be introduced in section 3 and 4. Conclusions and additional thoughts will be given in section 5.

## **2. Data Summary**

To address the two research questions, the response value we interest in was the improve rate; in other words, all measurements would be done for users who did the test more than once with more than one BLR score returned. The 17 characteristics used to measure in the BLR platform are: heart score; respiratory score; gastrointestinal score; diabetes score; cancer score; arthritis pain score; mental health score; social financial relationship score; diet score; physical activity score; financial health score; medication score; alcohol score; sleep score; stress score; smoking score and BMI. Also, user's birth year; gender; created date and finished date for one test were recorded. A higher BLR score means healthier.

The aim of manipulation was to create a new data set containing the change ratio of each score over the period of doing test per person. Time period recorded in the new data set would be the difference of last BLR finished date and the first BLR finished date for each user.

Steps for first question manipulation is below:

- Restricted users whose BMI value is between 13 and 40; others were ignored;
- Time length to finish each test wasn't longer than 14 days; others were ignored;
- Arranged the dataset with a descent order of the finished date within each user;
- Appended the change ratio of BLR score for each user;

- Appended time difference of first finished date and last finished date for each user; filtered observations if the time difference is 0;
- Appended change ratio for each characteristic score per user: heart score ratio; respiratory score ratio; gastrointestinal score ratio; diabetes score ratio; cancer score ratio; arthritis pain score; mental health score; social financial relationship score ratio; diet score ratio; physical activity score ratio; financial health score ratio; medication score ratio; alcohol score ratio; sleep score ratio; stress score ratio; smoking score ratio; BMI ratio. Calculation were done according to the formula:  $change\_ratio = \frac{lastscore}{firstscore}$ .

The second question is going to find the relationship between frequency of usage and BLR score change rate. Based on manipulated dataset mentioned above, we calculated the times that each user login on the platform during the period of first created date and last finished date.

Steps for the second question manipulation are below:

- Restricted Z company's user points recorders between the time period of first BLR created date and last BLR finished date;
- Calculated the frequency of points records for each user from the company Z;
- Appended the user's frequency of usage on the new dataset created in the first question;
- Repeated for all company M and company R user, filtered observations whose frequency is 0.

There were too many missed value in the variable “Adherent level”, this is not considered in this analysis. A summary for the new manipulated dateset with score change ratio was shown in the Appendix.

### 3. Methods

Before fitting to the model, the response value and covariates values were calculated respictively. Each score ratio times the power of inverse of time difference. For example, the BLR change rate was calculated according to  $BLRchangeRate = (\frac{lastsBLR}{firstBLR})^{(1/timediff)}$ .

Simple linear regression model was used to address both questions. Analysis was carried out by using lm function in R. The first question considered the relationship between the BLR change rate and other characteristics. The model with “Gender” and “Birth Year” were considered, but neither of them were found to have sufficient influences on the BLR score. Therefore, the model adopted here includes other 17 characteristics conditions: heart; respiratory; gastrointestinal; diabetes; cancer; arthritis pain; mental health; social financial relationship; diet; physical activity; financial health; medication; alcohol; sleep; stress; smoking and BMI. Assumption for this question was that change rate of these characteristics do not have influences on the rate of BLR change. Second question analysis considered the relationship of BLR score change rate and frequency of usage. Also, birth year and gender were considered for further investigation. Hypothesis for this question was that frequency of usage was not associated to the general of BLR.

### 4. Results

Table1 shows the result to answer the first question on addressing the relationship of BLR improve rate and characteristics. There were sufficient evidences to show that all characteristics have influence on the BLR change rate with very small P-value. Average BLR change rate for all survey respondents is positive, which means that there was an 0.39 improvement from the first BLR test to the last BLR test on average. Characteristics scores of heart, respiratory, mental health, diet, physical activity, financial health, medication, alcohol, sleep and smoking have positive relationship with the BLR change rate; others had negative influence. In other words, one unit increase in the change rate of heart score; respiratory score; mental health score; diet score; physical activity score; financial health score; medication score; alcohol score; sleep score; stress score and smoking score would lead to 0.0024; 0.0015; 0.0015; 0.1053; 0.1542; 0.1232; 0.0076; 0.0055; 0.0830; 0.1254 and 0.1540 increase in the BLR change rate respectively; on the other side, one unit increase in the change rate of gastrointestinal score; diabetes score; cancer score; arithmetic score; social financial relationship score

and BMI score will lead to a decrease of 0.0021; 0.0304; 0.0078; 0.0062; 0.0055 in the BLR change rate respectively. Therefore, the characteristic of heart score; respiratory score; mental health score; diet score; physical activity score; financial health score; medication score; alcohol score; sleep score; stress score and smoking score would have impacts on the BLR score improvement. Some demonstration plots have shwon in the Figure3 in appendix section.

Table 2 indicates the results to interpret the relationship between the BLR change rate and frequency of usage. There was no sufficient evidence to show that frequency of usage has influence on the BLR change rate ( $p= 0.7775$ ). Figure1 has shown the relationship of BLR improvate rate and usage frequency, there is no obvious indication that higher frequency will lead to a higher BLR improve rate. A summary table for frequency value is attatched in the appendix.

To further investigate other covariates' influences on the BLR score change rate, we have carried out the analysis of other covariates in the model. Still, there is no sufficient evidence to show the birth rate have influence on the BLR change rate( $p=0.6540$ ). However, the factor of gender may have significant evidence to show influence on the BLR change rate with p-value = 0.0224. In other words, there was not too much differences on the BLR change rate for different ages of people; but the gender does. In specific, females had 1.0045 unit increase in the BLR improve rate compare to males. Figure2 has demonstrated the comparison between females and males for BLR improvement rate.Thus, it could not conclude that higher frequency of usage will lead to the increase of BLR change rate. Also, cannot conclude for the age difference on the BLR change rate; but females may had higher BLR improve rate than males.

Table 1: Summary Table for first question, indicates the relationship of BLR improve rate and other characteristics.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.3911	0.0079	49.2659	0.000
Heart.impro_stan	0.0024	0.0001	17.1282	0.000
Res.impro_stan	0.0015	0.0002	6.9416	0.000
Gas.impro_stan	-0.0021	0.0003	-8.1545	0.000
Diabetes.impro_stan	-0.0305	0.0008	-36.3455	0.000
Cancer.impro_stan	-0.0078	0.0002	-32.3829	0.000
Arth.impro_stan	-0.0062	0.0002	-27.9504	0.000
MentalHealth.impro_stan	0.0015	0.0002	8.7531	0.000
SocFinRela.impro_stan	-0.0055	0.0002	-25.1605	0.000
Diet.impro_stan	0.1053	0.0021	51.0372	0.000
PhyAct.impro_stan	0.1542	0.0021	74.5031	0.000
FinHeal.impro_stan	0.1232	0.0018	70.0186	0.000
Medic.impro_stan	0.0076	0.0007	11.2019	0.000
Alco.impro_stan	0.0055	0.0021	2.6541	0.008
Sleep.impro_stan	0.0830	0.0039	21.1855	0.000
Stress.impro_stan	0.1254	0.0026	48.5583	0.000
Smoking.impro_stan	0.1540	0.0027	57.1739	0.000
BMI.impro_stan	-0.1027	0.0059	-17.4172	0.000

Table 2: Summary Table for second question

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.0045	0.0116	86.9007	0.0000
Frequency	0.0000	0.0000	0.2826	0.7775
YearOfBirth	0.0000	0.0000	-0.4482	0.6540
Gendermale	0.0004	0.0002	2.2846	0.0224

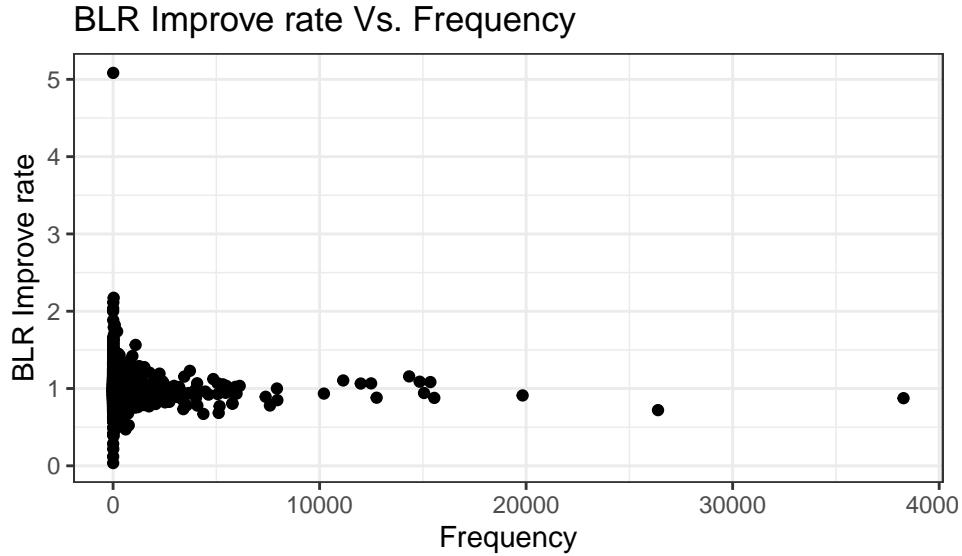


Figure 1: Plot of relationship between BLR Improve rate and frequency, more points were concentrated at small values. It does not show strong relationship.

## 5. Conclusions and Discussion

This analysis was mainly focusing on the relationship of BLR improve rate and other covariates. Characteristics of heart score, respiratory score, mental health score, diet score, physical activity score, financial health score, medication score, sleep score, stress score and smoking score may have influence on the BLR improvement rate. However, the frequency of usage of the platform didn't have relationship of the BLR improve rate. Further study showed that gender may have influence on BLR improve rate, but age does not.

Things to note in this study was that there were lots of observations neglected since the scores that engaging people reported were out of considerations. For example, some of BMI scores in the raw data were extremely large ( $>100.00$ ) or extremely small ( $<1.00$ ). Also, many user's period of test was too long or repeated, and some user did the tests twice or three time in one day. This might due to that engaging people were not doing the tests according to instructions or they recorded it by mistakes. Thus, these inaccurate records from engaging people might have influences on the accuracy of this analysis.

Sample size for this analysis was large enough, but more accuracy of recording would improve the analysis results.

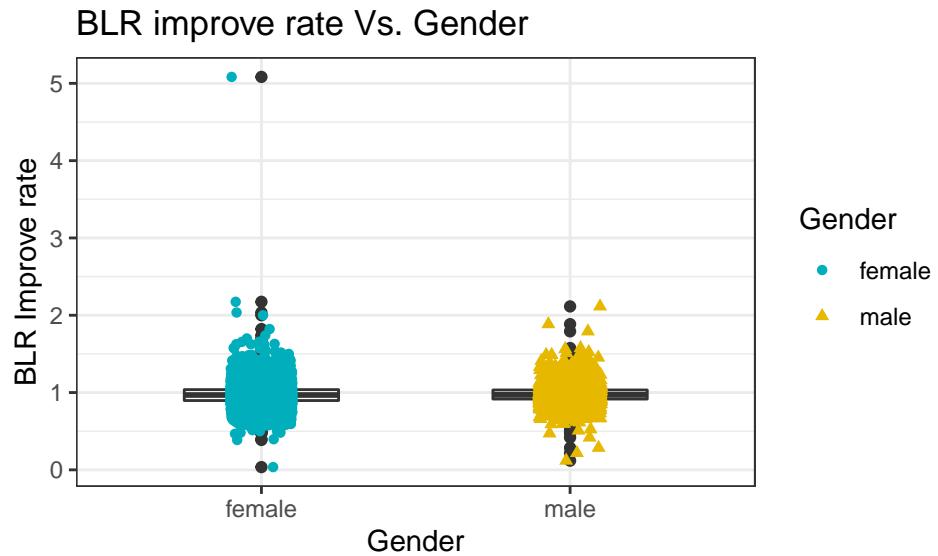


Figure 2: Plot of relationship and gender. The mean value were the same for females and males, but females had wider distribution, and the BLR improve rate is little higher than the males

## APPENDIX

Table 3: summary table for some variables

BLR.impro	timediff	Heart.impro	Res.impro	Gas.impro	Diabet.impro
Min. :0.03571	Min. : 1.0	Min. : 0.03333	Min. : 0.07143	Min. : 0.05882	Min. : 0.1000
1st Qu.:0.90476	1st Qu.: 183.0	1st Qu.: 0.92857	1st Qu.: 1.00000	1st Qu.: 0.93750	1st Qu.: 0.9167
Median :0.96825	Median : 363.0	Median : 1.00000	Median : 1.00000	Median : 1.00000	Median : 1.0000
Mean :0.97201	Mean : 444.8	Mean : 1.94790	Mean : 1.65041	Mean : 1.65558	Mean : 1.2048
3rd Qu.:1.03333	3rd Qu.: 617.0	3rd Qu.: 1.11539	3rd Qu.: 1.00000	3rd Qu.: 1.16667	3rd Qu.: 1.1429
Max. :5.08333	Max. :1649.0	Max. :30.00000	Max. :13.00000	Max. :18.00000	Max. :11.0000

Table 4: Summary table for usage frequency

new1.points\$Frequency
Min. : 2.0
1st Qu.: 4.0
Median : 11.0
Mean : 135.7
3rd Qu.: 51.0
Max. :38274.0

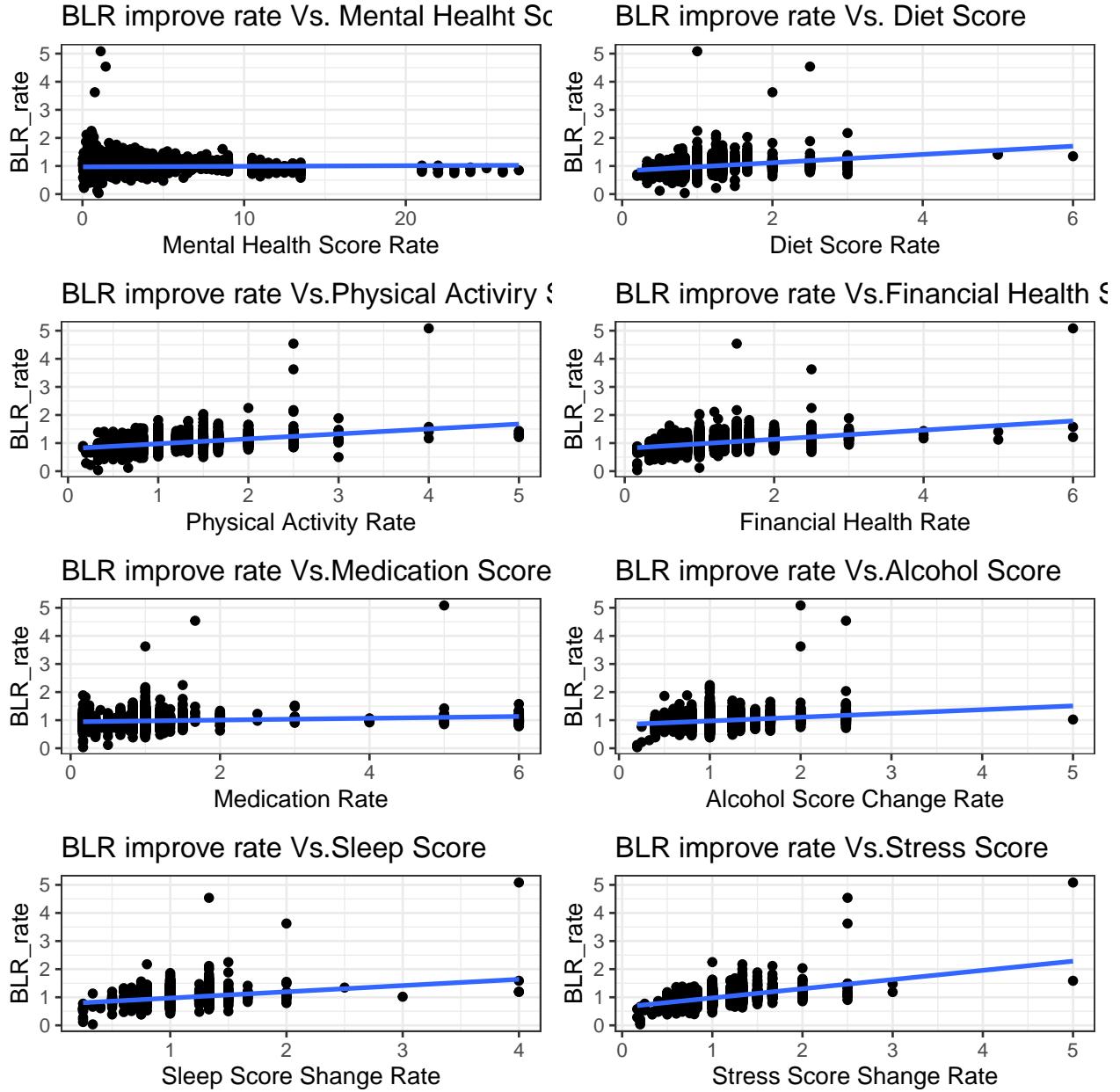


Figure 3: A combination of example plots shows the relationships between the BLR change rate and characteristics. These plots show a general positive relationship between characteristics and BLR improve rate.