Impact of improvement in user's characterstics and usage of BestLifeReward platform on BestLifeReward health score improvement

Yuhan Hu

2020/04/05

Abstract

While online health assessment apps are getting popular, their effectiveness can be improved and usefulness requires further investigation. This report was based on data collected from the platform developed by BestLifeReward, which is an online health assessment platform for company emplyees. Primarily, We want to investigate improvement in what characteristics of users had greatest impact on rate of change of health status, which is represented by an overall health score, BLR score, in this analysis. Furthermore, we are also interested in association between platform usage frequency and health improvement. Here we report that improvement in physical activity score, diet score, financial health score, stress score, smoking score and medication score are positively related to BLR score improvement. In addition, there is no evidence showing higher platform usage frequency are related to general health improvement.

Introduction

An increasing number of individuals are using online platform for health guidance and assessment nowadays. By utilizing these tools, users may evaluate personal health status by recording health-related aspects, and get guidance for living healthier. However, quality of these platform require further study.(Grundy, Wang, and Bero 2016)

In this report, we focused on the online platform for employee health assessment provided by BestLifeRewarded Innvations. This platform collaborate with employers to provide resources and programs which helps employees manage their health status. On this platform, users can evaluating personal health status by completing survey on aspects like individual activities(e.g. physical activities) and health-related factors(e.g. weight and height). Based on survey results, the platform will calculate an overall health score, BLR score, and provide guidance for healthy life behaviour for users. Also, in order to motivate usage of platform, users can earn reward points when participating in platform activities. We were primarily interested in what characteristics of users were associated with higher rate of overall health improvement. Or in another word, how does changes in those characteristics affect BLR score rate of change. Answer to this question can help the platform provide better service by focusing more on those characteristics. In addition, we also investigated whether a higher frequency of platform usage would lead to a higher rate of overall health improvement for the sake of proving usefulness of the platform.

This report is stuctured to present analysis process and result about these two research question as well as discussion about limitations and future directions. In method section, we discussed about how the platform works, data summary and manipulation as well as statiscal analysis. In result section, result of the study are provided, illustrated by tables and plots. Conclusion of major findings, potential question underlying this analysis and possible direction for future research are in discussion section.

Method

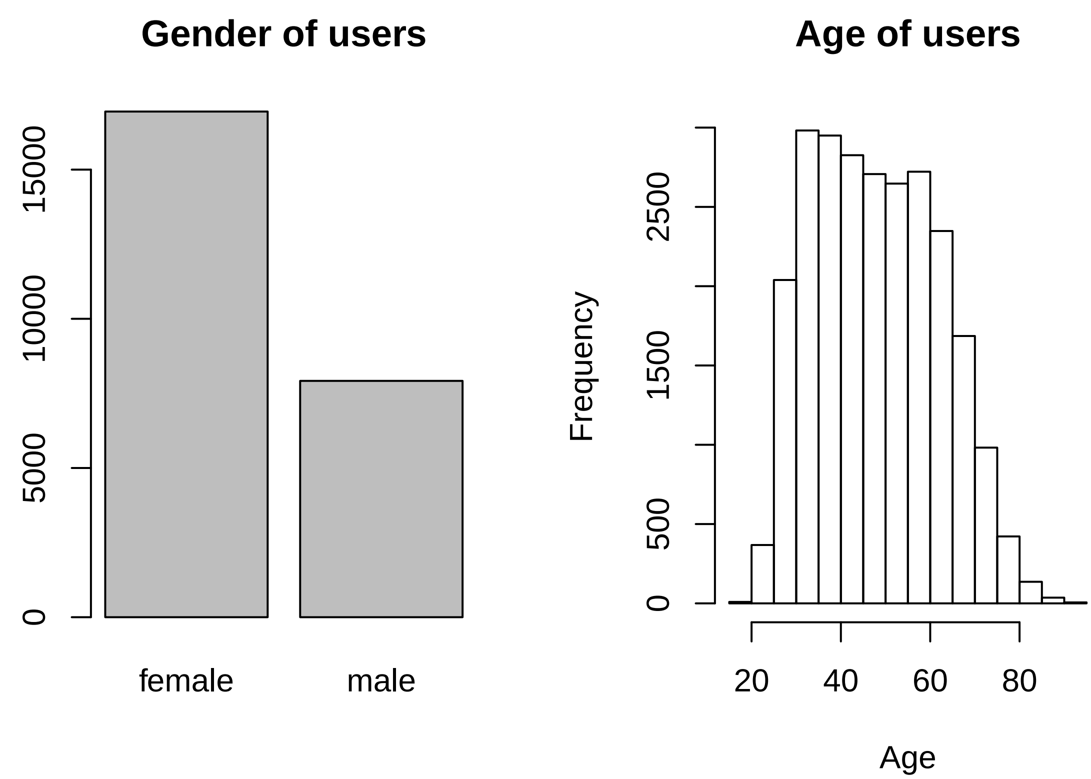
Data collection, summary and Manipulation

Data used for this study was collected from users on BestLifeReward platform. All participants came from one of three companies. Based on informaiton collected from users, such as individual activities(sleeping, nutrition, physical activities, etc.) and health-related factors(height, weight, stress level, etc.), scores are evaluated for health-related aspects such as gastrointestinal score, cancer score, mental health score, diet score, physical activity score, financial health score, medication score, alcohol score, stress score, smoking score, heart score, respiratory score, diabetes score, arthritis pain score, Social financial relationship score, sleep score and BMI. An overall health score, BLR scores will be calculated based on those health risk factors.

In our original dataset, there are 170965 data points in total. 32197 entries are removed because they have N/A for at least one feature. User can alter data entries after created. Dates of creation and last alteration were recorded as created date and finished date, respectively. However, if a data entry was altered several days after created, this data is unlikely to reflect a concise health state of the user. Therefore, we remove all data points whose created date is not the same as finished date. At this step, additional 15370 data points are removed.

Recall the first research question, chronological changes between data points from same user was the data of interest. So we did not use original value for analysis. Instead, for each user, we subtracted values in the first entry by values in the last entry(chronological order), then divided the result by time difference between this two entry. In addition, if a user only had one valid entry, such subtraction can not be performed. Therefore, we removed users with only one valid entry before subtraction and normalization. Additional 83409 data entries are removed at this step. After remove all this data points, there are 72186 entries left. In addition we found there are users who input multiple entries on one day. Validity of these entries are questionable; thus we removed those data. 2702 entries removed at this step. There were 69484 original data points from

24866 unique users after cleaning. This complete data preparation for linear regression model of research question 1. Among these users, about 2/3 are female, and most of them are 30-60 years old.



We also built a dataset for linear mixed model of research quesiton 1. Instead of substracting first entry from last entry, here we subtract second last entry from last entry, then third last entry from second last entry, all the way to the first entry.

For the second research question, since **frequency of platform usage** is not clearly defined by the collaborator, we obtain several different dataset based on distince definition of **frequency of platform usage**:

1. Represent frequency of platform usage as sum of accumulation of points. i.e. added up total points earned for each user
2. Represent frequency of platform usage as sum of points spent. i.e. added up total points spent for each user
3. Represent frequency of platform usage as absolute change in points. i.e. add absolute value of total points spent to total points earned.

4.Represent frequency of platform usage as ratio of sum of accumulaiton of points to number of platform usage. i.e. divided total points earned by number of platform usage.

What characteristics are related to higher rate of BLR score change?

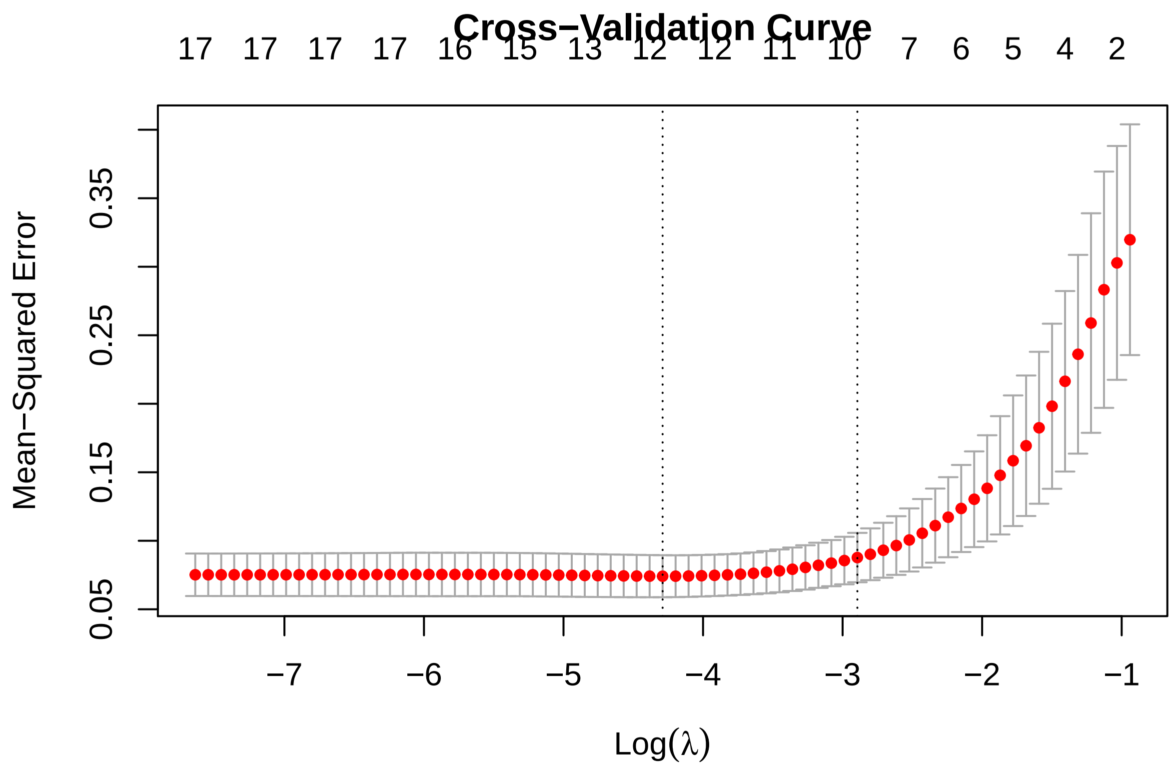
In our data, one participant may contribute to multiple data points. Apparently, observations from same participant are not independent. In addition, homoscedasticity assumption of linear regression may also be violated. Therefore, we built a linear mixed model first. However, based on result from linear mixed model, percentage of variability explained by random effect are too small to be substantial.(i.e. given 3 digits, $\frac{\sigma^2}{\sigma^2+\tau^2} = 0$)Therefore, we decided linear mixed model was overcomplicating for this study and not supposed to be used here.

Table 1: Linear Mixed Model for Research Question 1

	MLE	Std.Error	DF	t-value	p-value
(Intercept)	0.082	0.056	29486	1.470	0.141
Heart_change	-0.084	0.010	29486	-8.060	0.000
Respiratory_change	-0.113	0.014	29486	-7.970	0.000
Gastrointestinal_change	-0.203	0.014	29486	-14.693	0.000
Diabetes_change	-0.170	0.018	29486	-9.422	0.000
Cancer_change	-0.428	0.015	29486	-27.886	0.000
MentalHealth_change	-0.274	0.012	29486	-22.804	0.000
SocioFinancial_change	-0.283	0.017	29486	-16.641	0.000
Diet_change	1.419	0.022	29486	65.851	0.000
PhysicalActivity_change	1.416	0.019	29486	73.201	0.000
FinancialHealth_change	1.267	0.017	29486	76.325	0.000
Medication_change	0.324	0.018	29486	18.494	0.000
Alcohol_change	1.226	0.031	29486	39.359	0.000
Sleep_change	1.182	0.040	29486	29.879	0.000
Stress_change	1.198	0.030	29486	40.614	0.000
Smoking_change	0.812	0.013	29486	64.353	0.000
Gendermale	-0.044	0.032	29486	-1.365	0.172
Age	0.001	0.001	29486	1.346	0.178
σ	0.000	NA	NA	NA	NA
τ	2.926	NA	NA	NA	NA

Changes in BLR score is the variable of interest, thus the response variable. All other variables are potential predictor of interest. Since we do not have additional information about covariates, there is no clue for selecting significant features requried by our model. Therefore, we decide to use LASSO shrinkage method to choose features automatically. Since we do not have additional information about possible interaction between potential variables, no interaction terms are added into the model.

As for tunning parameter λ of LASSO shrinkage, we chose λ_{1se}(second vertical dotted line) because this λ value gives a model that error is within one standard error of the minimum, which is the most regularized model. (Friedman, Hastie, and Tibshirani 2010, glmnet package vignette)



By LASSO shrinkage, changes in gastrointestinal score, cancer score, mental health score, diet score, physical activity score, financial health score, medication score, alcohol score, stress score and smoking score are included in the model, while changes in heart score, respiratory score, diabetes score, arthritis pain score, social financial relationship score, sleep score and BMI are removed from the model.

Is a higher frequency of usage of the platform associated with a general improvement in health as measured by the risk factors?

As for this research question, since the idea of **frequency** was not clearly defined, we have several potential predictors such as sum of points earned, sum of points spent, times of platform usage and ratio of sum of points earned to times of platform usage. Response variable was the same as previous research question, change in BLR scores.

Linear regression models has been fitted for each predictors individually, and only ratio of sum of points earning to times of platform usage returns a coefficient with significant p-value.

Table 2: model with ratio of sum of points earning to times of platform usage as predictor

	coefficient
(Intercept)	2.6411799
points_per_use	-0.0000025

Table 3: model with total points earned as predictor

	coefficient
(Intercept)	2.640704
points_earned	0.000000

Table 4: model with total point spent as predictor

	coefficient
(Intercept)	2.640709
points_used	0.000000

Table 5: model with absolute points change as predictor

	coefficient
(Intercept)	2.640708
sum_points	0.000000

Result

Research Question 1

As mentioned in method section, we are using LASSO regression to do model selection for this question. After shrinkage with apporiate λ, we found a final model with 11 predictors, which included changes in gastrointestinal score, cancer score, mental healthscore, diet score, physical activity score, financial health score, medication score, alcohol score, stress score and smoking score will stay in the model for inference. Adjusted R^2 is 0.808, which suggest that 80.8 of variation in BLR improvement was explained by this model.

Refer to the coefficient Table, changes in physical activity scores has the greatest coefficient, 2.156, which indicate that improvement in physical activity scores was associated with the highest improvement rate of BLR score. Or more specifically, increase physical activity score by 1 unit will increase BLR score by 2.156 units on average, given other predictors did not change. Improvement of alcohol score, medication score, smoking score, stress score, financial health score and diet score also had positive coefficient, which indicate improvement in these scores led to increase BLR score. Changes in cancer score, gastrointestinal score and mentalhealth score had negative coefficient. Therefore, the larger these 3 scores, the smaller the BLR score, which indicates a worse health condition. Among these 3 scores, changes in mental health score has the largest absolute value, 0.541, which indicate increase mental health score by 1 unit will reduce BLR score by 0.541 unit on average, given other predictors did not change.

Both negative effect predictors and positive predictors are given in increasing order of effect on BLR score change.(i.e. in increasing order of absolute value of corresponding coefficients.)

Table 6: Post-LASSO Coefficient of health-risk factors

	x
(Intercept)	0.003
GastrointestinalScore_change	-0.190
CancerScore_change	-0.344
MentalHealthScore_change	-0.498

	x
DietScore_change	1.572
PhysicalActivityScore_change	2.192
FinancialHealthScore_change	0.933
MedicationScore_change	0.736
AlcoholScore_change	1.050
StressScore_change	0.989
SmokingScore_change	1.224

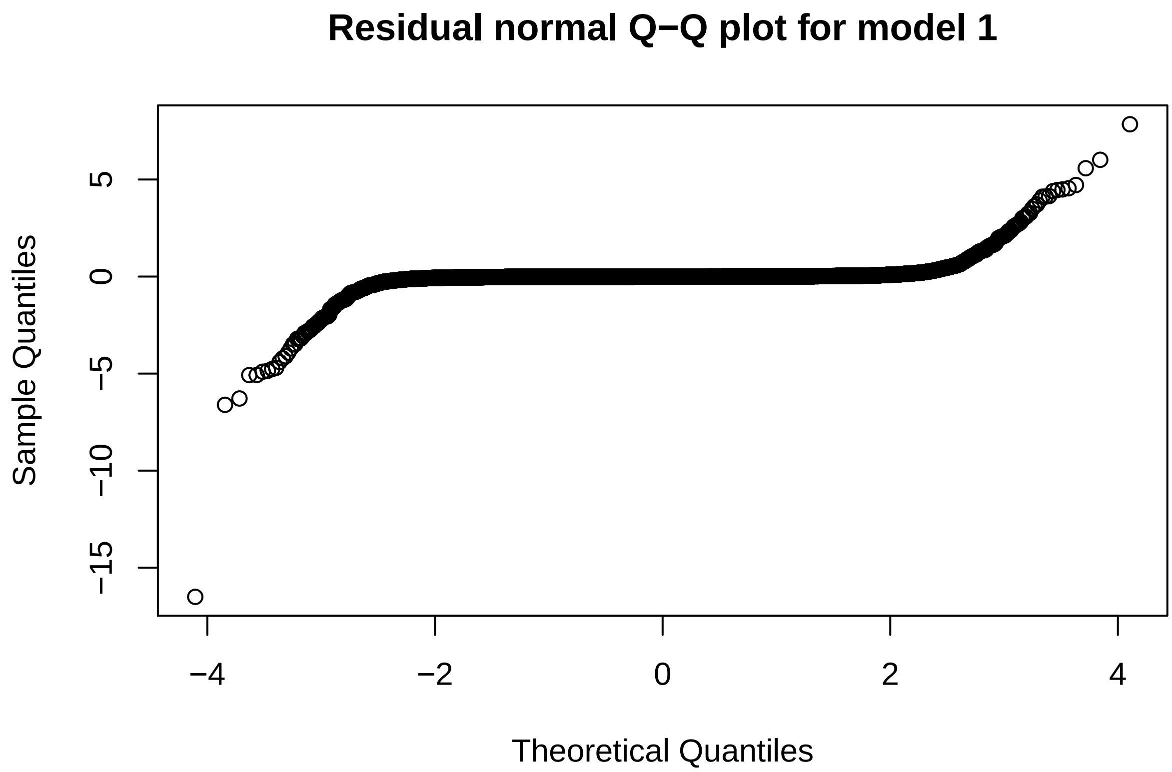
Research Question 2

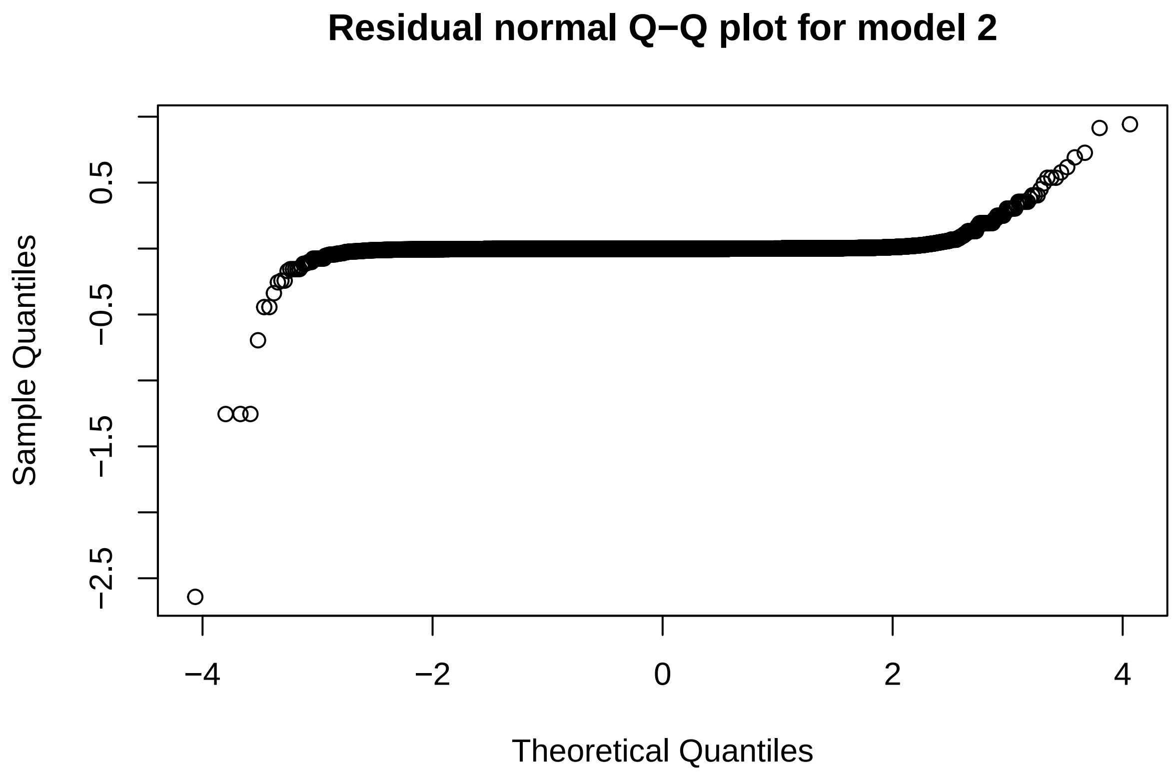
As mentioned and illustrated above, most potential predictors do not return a significant coefficient, which indicates that those predictors did not affect changes in BLR scores. Moreover, value of the only significant coefficient, coefficient of ratio of sum of points earning to times of platform usage, is close to 0, and the corresponding p-value is at the boundary between significant and non-significant. Therefore, the answer to the second research question is negative. Higher frequency of platform usage does not associate with a general improvement in BLR score.

Discussion

This analysis utlize LASSO shrinkage to select features for determining association between improvement in health risk factors and BLR score improvement. λ we chose value for LASSO led to the most regularized model whose error is within one standard error of the minimum. Result of post selection inference shows that BLR score improvement is positively related to alcohol score, medication score, smoking score, stress score, financial health score, diet score and physical activity score. The improvement of users' overall health status is negatively associated with increasing cancer score, gastrointestinal score and mentalhealth score. Both positive and negative effects were listed in increasing absolute value order. As for relationship between platform usage frequency and BLR score improvement, we attempted several linear model with various definition of platform usage frequency, and obtained consistent conclusion from all models. Refer to our result, there was no evidence shows that higher platform usage frequency was associated with general improvment in BLR score.

Moreover, there are several aspects worth mentioning about this analysis. First and the most important, residual plot of both models are heavily tailed, we have tried various method like outlier removal and data transformation(e.g. log transformation), but the problem cannot be solved.


Residual normal Q−Q plot for model 1
Sample Quantiles
Theoretical Quantiles
5
0
−5
−10
−15
−4
−2
0
2
4



Model-specific discussion are given below.

Model 1

1. As we have mentioned, variability explained by personal difference in linear mixed model is so small that can be ignored. For furture analysis, this result can be helpful during model selection.
2. Process of predictor selection was done by LASSO algorithm automatically. However, in analysis, there might be variables that are supposed to be in the model regardless of measure of significance. For example, we can not present a model without predictor of interest even if the this variable had a non-significant p-value. Or another example, a necessary control variable may have a non-significant p-value, but it can not be removed. Due to lack of information, we didn't know if there were such variable.
3. Variance inflcation factor for changes in gastrointestinal scores and mental health scores in this model are greater than 3. Though we decided to ignore that because the most commonly used standard for VIF is rule of 10 (O'brien 2007), exactly how large a VIF has to be before it causes issues is a subject of debate. So if furture analysts prefer a very strict standard, such as 2.5, they need to concern about multicolinarity here.

Table 7: VIF table

	VIF
GastrointestinalScore_change	3.224
CancerScore_change	1.246

	VIF
MentalHealthScore_change	3.113
DietScore_change	1.347
PhysicalActivityScore_change	1.536
FinancialHealthScore_change	1.375
MedicationScore_change	1.125
AlcoholScore_change	1.079
StressScore_change	1.739
SmokingScore_change	1.063

Model 2

1. The research question is ambiguous since **'frequency of usage of the platform'** was not clearly defined. There was no variable labeled as 'frequency' in original files provided by collaborator and the collaborator did not tell us which variable matched their idea of 'frequency of platform usage'. Moreover, we were not given information about how to represent **'frequency of usage of the platform'** using data from original files. Though we tried to fit several models with various representation of **'frequency of usage of the platform'** and obtained consistent conclusion, if collaborator's idea of **'frequency of usage of the platform'** was not in our set of definitions, result from current analysis may become invalid.
2. Refer to our result, higher frequency of platform usage is not associated with higher rate of BLR score improvement, which is counter-intuitive. Intuitively speaking, people were supposed to perform better with help from supervisor and advisor. Reason underlying this counter-intuitive result can be a direction for future investigation.

Suggestion for furture data collection

At last, we have some suggestions for data collection process. As we have mentioned in data manipulation part, the original data files contain too many invalid data entries,which is evitible if the data collection process was refined. Specific suggestion are listed below:

1. Add time limit for data entry. An data entry that are altered two weeks after creation cannot reflect a concise health state.
2. Creating multiple entries in one day are not supposed to be allowed. Users are supposed to alter the created entry rather than create a new one.
3. Limit range for variables. There were non-human data input in the original data files. For example, there were observations with BMI value less than 1.

References

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1.

Grundy, Quinn H, Zhicheng Wang, and Lisa A Bero. 2016. "Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods." *American Journal of Preventive Medicine* 51 (6): 1051–9.

O'brien, Robert M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality & Quantity* 41 (5): 673–90.