
REPORT

NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY (NHANES) 2013-2014

1.0 BUSINESS UNDERSTANDING

1.1 Problem Description

An estimated 790,000 United States (US) adults suffer from myocardial infarction also known as heart attack each year which suddenly denied oxygenated blood from reaching the heart muscle. (Elizabeth, 2017) If the condition is not treated promptly, permanent heart damage or even death ensues.

However, one in every five heart attack hits the victim when they least expected, which suggests that they often are not rendered with proper treatments and had acted on early warning risk signs. (Benjamin, Blaha, Chiuve, Cushman, Das and Deo, 2017)

Therefore, the team intend to develop a model that can identify the prevalent factors that can successfully predict a heart attack so that the authority can pick up the subjects at risk and develop programs that can mitigate the risk conditions.

The targeted outcome is to reduce the heart attack rate of US adults by 10% in 10 years.

1.2 Business Objective and Goals

The team will be using the self-reported datasets from the National Health and Nutrition Examination Survey (NHANES) which provides a unique statistical representation of the country population's personal information, general health, habits and diet. (NCHS, 2018)

We hope to extract information to gain insights on the tell signs of the heart attack patients that would be usefully for us to predict who are at risk of suffering from a heart attack.

1.3 Methodology

The team will be using Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to systematically structure the data mining processes. It consists of six phrase which are Business Understand, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment.

As the selected datasets are complex and may have nonlinear relationships between dependent and independent variables, we will be employing Neural Networks (NN) to model the problem domain. Since our objective is the prevention and not the diagnosis of heart attack, we can mitigate the disadvantages of the "Black Box" nature mentioned by Tu (1996) that plagued neural network model adoption in predicting medical outcome.

The R package CARET (Classification And REgression Training) provided the functions to streamline the neural network model training, testing and building of the ensemble.

We will train a single layer perceptron model from NNET library, multi-layer perceptron model from MXNET library and radial basis function with dynamic decay adjustment algorithm model from RSNNS library in R.

Lastly, we will utilise the caretStack function from the caretEnsemble library with K-nearest neighbour algorithm in R to create the ensemble of the three NN model for the final predication.

2.0 DATA UNDERSTANDING

2.1 Data Collection

The team selected the National Health and Nutrition Examination Survey (NHANES) datasets collected by the National Center for Health Statistics (NCHS) of the United States which covered a comprehensive aspect of the population's health and nutrition status.

The survey was done in the period between 2013 to 2014 which is consider up to date for a more accurate interpretation of the model results when developing policies.

2.2 Data Exploration

The National Health and Nutrition Examination Survey (NHANES) datasets of 6 datasets the are as follows: -

DATASET_NAME	VARIABLES	OBSERVATIONS
Demographics	47	10175
Examinations	224	9813
Dietary	168	9813
Laboratory	424	9813
Questionnaire	953	10175
Medications	13	10175

The dataset name is self-explanatory of the content nature of the datasets. Respondent sequence number are common in all six datasets which allow us to logically and relationally join up tables. The resultant dataset consists of 1829 of both numeric and categoric variables. There are total of 10175 observations. The selected variable (IE. HeartAttack_MyocardialInfarction) had more than 5000 observations which the team felt was sufficient for our modelling.

2.3 Data Verification

The NHANES datasets were collected by the National Center for Health Statistics (NCHS) which is a research oriented statistical organisation. The survey is commissioned by Department of Health and Human Services of the United States, a national body. Therefore, the survey was professionally done with no influence by lobbying groups.

In addition, the self – reporting nature of the survey was reinforced with medical and laboratory testing data conducted by professionals which adds up to the reliability of the datasets.

3.0 DATA PREPARATION

3.1 Data Selection

The merged dataset was further reduced by firstly removing variables that are irrelevant to our objective through research on the medical terms and heuristically. Secondly, variables with less than 5000 observations which deem not related to the target variable are dropped. Lastly, missing response form the target variable are removed.

The final based dataset had 86 variables and 5585 observations which is more manageable and relevant to the problem domain.

3.2 Features Selection

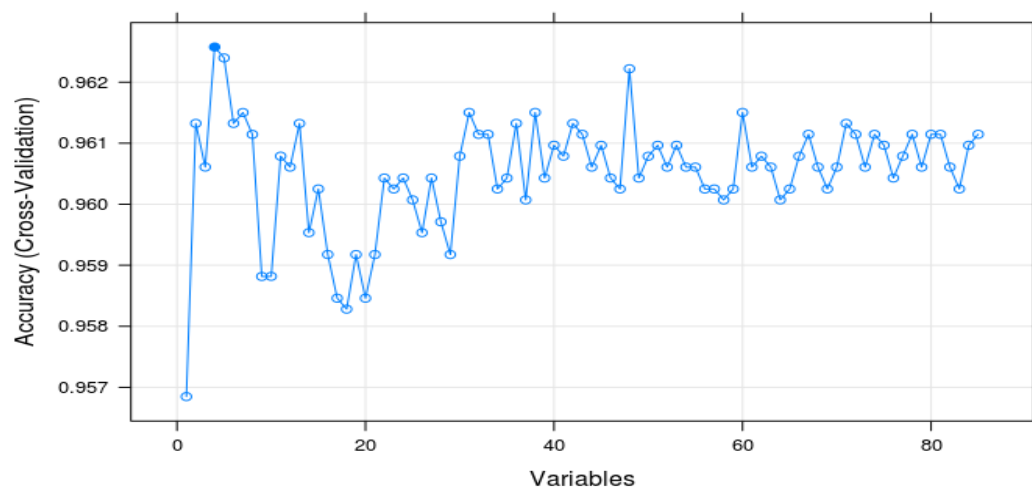
The team attempted two techniques, Recursive Feature Elimination and Correlation method for feature selection to further reduce the number of variables. We hope to be able to find the most appropriate method to help us identify the most significant variables.

3.2.1 Recursive Feature Elimination

The Recursive Feature Elimination function in Caret R package provides the tool to automatically report on the relevance and importance of attributes in the base dataset. The details of the runs were as follows: -

Resampling performance over subset size: 85

Outer resampling method: Cross-Validated (5 fold)



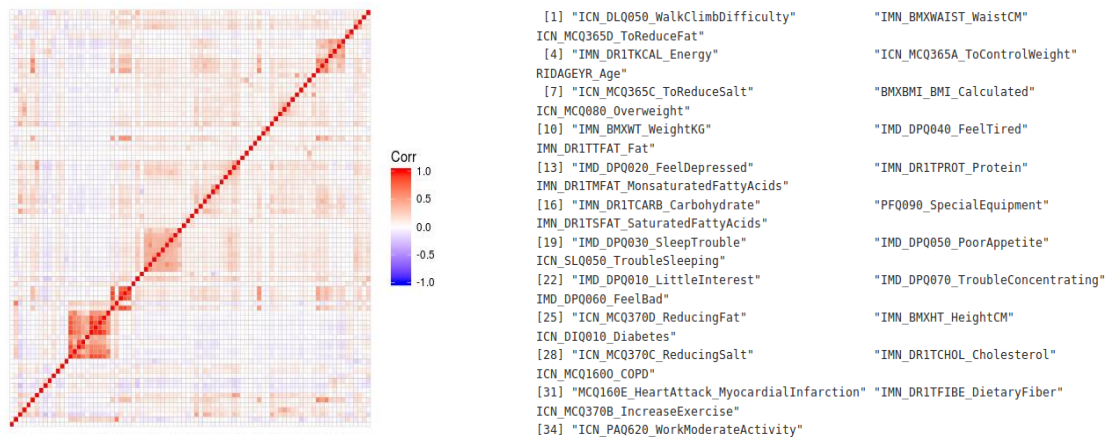
The top 4 variables (out of 4):

ICN_MCQ160C_CoronaryHeartDisease, ICN_MCQ160B_CongestiveHeartFailure, ICN_MCQ160D_Angina, ICN_RXQ510_TakingAspirinForHeartAttack

From the above output, we can see that a model size of the top four variables seems to achieve the optimal accuracy and out performed a larger feature size model. But if we only choose 4 out of 85 variables we may leave out too much information that could help the authority formulate policies. Therefore, we did not use the results for feature selection.

3.2.2 Correlation Matrix

The team generated a correlation matrix and selected variables that are 0.4 correlated. The resultant dataset reduced from 86 to 34 variables which is fair reasonable reduction.



3.3 Data Scaling and Partitioning

The team split the original and correlated features datasets into 80 % for training and 20% for testing using the R package call DMWR. The training datasets minority were oversampled to 10% and majority were also oversampled so that each training dataset will reach 10`000 total observations with the ROSE package in R.

The transformation of the data by scaling (IE. Divide by standard deviation) and center (IE. Subtract Mean) are handled by the train function pre-processing in CARET and will be part of the output model so that each variable will have the same input range in prediction during the training.

We tested ran the training original and correlated features datasets on a multi-layer perceptron in MXNET. The result showed that that correlated features model did not perform better than the model with the original dataset.

Therefore, in the subsequence model building we will be using all the original dataset features (IE. 86 variables).

4.0 MODELLING

The team will be using caret package in R as a framework for our model development. It integrates and streamlined the workflows of all the related activities in the model developments. We have selected 3 models that have difference architecture so that the final assembled ensemble will be comprehensive and representative.

The 3 models are as follows: -

RELATED_MODEL_PACKAGE	NEURAL_NETWORK_TYPE
nnet	Single hidden layer feed forward perceptron
mxnet	Multi-layered feed forward perceptron
rbfDDA	Radial Basis Function with Dynamic Decay Adjustment Algorithm

4.1 Building Model

The training control settings were defined as k-fold cross validation with k = 10, saving class probability and prediction were set as true. This training setting will be constant for all the model trained.

4.1.1 Single Hidden Layer Feedforward Perceptron (nnet)

The R package NNET provides methods for using feed-forward neural networks with a single hidden layer. It is a neural network where connections between the units do not form a directed cycle

The following are the tuning parameters provide by caret for NNET: -

model	parameter	label	forReg	forClass	probModel
<chr>	<fctr>	<fctr>	<lgl>	<lgl>	<lgl>
1 nnet	size	#Hidden Units	TRUE	TRUE	TRUE
2 nnet	decay	Weight Decay	TRUE	TRUE	TRUE
2 rows					

The above diagram shows that the nnet have two parameters and is suitable for classification problem. The best fit tune parameters the team found were size = 3 and decay = 0.9.

The following is the confusion matrix of the nnet for training dataset: -

Confusion Matrix and Statistics

```
Reference
Prediction No Yes
No 7787 758
Yes 227 1228

Accuracy : 0.9015
95% CI : (0.8955, 0.9073)
No Information Rate : 0.8014
P-Value [Acc > NIR] : < 2.2e-16

Sensitivity : 0.6183
Specificity : 0.9717
Pos Pred Value : 0.8440
Neg Pred Value : 0.9113
Prevalence : 0.1986
Detection Rate : 0.1228
Detection Prevalence : 0.1455
Balanced Accuracy : 0.7950

Kappa : 0.656
McNemar's Test P-Value : < 2.2e-16
'Positive' Class : Yes
```

The following is the confusion matrix of the nnet for testing dataset: -

Confusion Matrix and Statistics

```

      Reference
Prediction No  Yes
No      1049  19
Yes      23   25

      Accuracy : 0.9624
      95% CI : (0.9495, 0.9727)
No Information Rate : 0.9606
P-Value [Acc > NIR] : 0.4175

      Sensitivity : 0.56818
      Specificity : 0.97854
      Pos Pred Value : 0.52083
      Neg Pred Value : 0.98221
      Prevalence : 0.03943
      Detection Rate : 0.02240
      Detection Prevalence : 0.04301
      Balanced Accuracy : 0.77336

      Kappa : 0.5239
McNemar's Test P-Value : 0.6434

      'Positive' Class : Yes

```

4.1.3 Radial Basis Function with Dynamic Decay Adjustment (rbfDDA)

Radial basis function (RBF) network with dynamic decay adjustment (DDA) algorithm is a feed forward network with one hidden layer that have an activation that is radially symmetric which is often in gaussian. The R package RSNNS provides method for the implementation of such architecture.

The dimensionality of the input space is represented by the number of input layer units. The hidden layer contains the RBF units which are added in during training. Each hidden unit is connected to exactly one output unit. The output layer units represent one possible class of a binary coding which is based on the winner takes all approach.

The following is the tuning parameter provide by caret for rbfDDA: -

model	parameter	label	forReg
<chr>	<fctr>	<fctr>	<lgl>
1 rbfDDA	negativeThreshold	Activation Limit for Conflicting Classes	TRUE

1 row | 1-5 of 6 columns

The above diagram shows that the nnet have one parameter and is suitable for classification problem. The best fit tune parameters the team found was negativeThreshold = 0.00000000000001.

The following is the confusion matrix of the rbfDDA for training dataset: -

Confusion Matrix and Statistics

```

      Reference
Prediction No  Yes
No      8014   0
Yes       0 1986

      Accuracy : 1
      95% CI : (0.9996, 1)
No Information Rate : 0.8014
P-Value [Acc > NIR] : < 2.2e-16

      Sensitivity : 1.0000
      Specificity : 1.0000
      Pos Pred Value : 1.0000
      Neg Pred Value : 1.0000
      Prevalence : 0.1986
      Detection Rate : 0.1986
      Detection Prevalence : 0.1986
      Balanced Accuracy : 1.0000

      Kappa : 1
McNemar's Test P-Value : NA

      'Positive' Class : Yes

```

The following is the confusion matrix of the rbfDDA for testing dataset: -

Confusion Matrix and Statistics

```

      Reference
Prediction  No  Yes
      No 1067  38
      Yes   5   6

      Accuracy : 0.9615
      95% CI : (0.9484, 0.972)
      No Information Rate : 0.9606
      P-Value [Acc > NIR] : 0.4787

      Kappa : 0.2057
      Mcnemar's Test P-Value : 1.061e-06

      Sensitivity : 0.136364
      Specificity : 0.995336
      Pos Pred Value : 0.545455
      Neg Pred Value : 0.965611
      Prevalence : 0.039427
      Detection Rate : 0.005376
      Detection Prevalence : 0.009857
      Balanced Accuracy : 0.565850

      'Positive' Class : Yes

```

4.1.3 Multi-Layer Feed Forward Perceptron(mxnet)

Multi-layer perceptron enables the learning of linear and non-linear separable patterns through by introducing at least one hidden layer to learn representation of the data that would enable linear separation.

The R package developed by Apache call MXNET provides the framework for the implementation.

The following is the tuning parameter provide by caret for mxnet: -

model	parameter	label	forReg	forClass	probModel
<chr>	<fctr>	<fctr>	<lgl>	<lgl>	<lgl>
1 mxnet	layer1	#Hidden Units in Layer 1	TRUE	TRUE	TRUE
2 mxnet	layer2	#Hidden Units in Layer 2	TRUE	TRUE	TRUE
3 mxnet	layer3	#Hidden Units in Layer 3	TRUE	TRUE	TRUE
4 mxnet	learning.rate	Learning Rate	TRUE	TRUE	TRUE
5 mxnet	momentum	Momentum	TRUE	TRUE	TRUE
6 mxnet	dropout	Dropout Rate	TRUE	TRUE	TRUE
7 mxnet	activation	Activation Function	TRUE	TRUE	TRUE
7 rows					

The above diagram shows that the mxnet have seven parameters and is suitable for classification problem. The best fit tune parameters the team found were layer1= 750, layer2=350, layer3=150, learning rate=0.005, momentum=0.98, dropout rate=0.000001 and activation function=relu.

The following is the confusion matrix of the mxnet for training dataset: -

Confusion Matrix and Statistics

		Reference		
Prediction	No	Yes		
No	8014	349		
Yes	0	1637		
Accuracy : 0.9651				Sensitivity : 0.8243
95% CI : (0.9613, 0.9686)				Specificity : 1.0000
No Information Rate : 0.8014				Pos Pred Value : 1.0000
P-Value [Acc > NIR] : < 2.2e-16				Neg Pred Value : 0.9583
				Prevalence : 0.1986
				Detection Rate : 0.1637
				Detection Prevalence : 0.1637
				Balanced Accuracy : 0.9121
Kappa : 0.8826				
McNemar's Test P-Value : < 2.2e-16				'Positive' Class : Yes

The following is the confusion matrix of the mxnet for testing dataset: -

Confusion Matrix and Statistics

		Reference		
Prediction	No	Yes		
No	1071	38		
Yes	1	6		
Accuracy : 0.9651				Sensitivity : 0.136364
95% CI : (0.9525, 0.975)				Specificity : 0.999067
No Information Rate : 0.9606				Pos Pred Value : 0.857143
P-Value [Acc > NIR] : 0.2483				Neg Pred Value : 0.965735
				Prevalence : 0.039427
				Detection Rate : 0.005376
				Detection Prevalence : 0.006272
				Balanced Accuracy : 0.567715
Kappa : 0.2269				
McNemar's Test P-Value : 8.185e-09				'Positive' Class : Yes

4.2 CaretStack Ensemble with K-Nearest Neighbour Algorithm

The `caretList` function uses the input of the optimized tuning parameters we found earlier for each of the model to build the caret model list with the same training data and training parameter control for consistency.

The following is the confusion matrix for the ensemble for testing dataset: -

Reference		
Prediction	No	Yes
No	1061	31
Yes	11	13
Accuracy : 0.9624		
95% CI : (0.9495, 0.9727)		
No Information Rate : 0.9606		
P-Value [Acc > NIR] : 0.41755		
Kappa : 0.3647		
McNemar's Test P-Value : 0.00337		
Sensitivity : 0.29545		
Specificity : 0.98974		
Pos Pred Value : 0.54167		
Neg Pred Value : 0.97161		
Prevalence : 0.03943		
Detection Rate : 0.01165		
Detection Prevalence : 0.02151		
Balanced Accuracy : 0.64260		
'Positive' Class : Yes		

5.0 EVALUATION

CONFUSION MATRIX - NNET – TEST

Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	1049	19
Yes	23	25

Accuracy : 0.9624
95% CI : (0.9495, 0.9727)
No Information Rate : 0.9606
P-Value [Acc > NIR] : 0.4175

Kappa : 0.5239
McNemar's Test P-Value : 0.6434

Sensitivity : 0.56818
Specificity : 0.97854
Pos Pred Value : 0.52083
Neg Pred Value : 0.98221
Prevalence : 0.03943
Detection Rate : 0.02240
Detection Prevalence : 0.04301
Balanced Accuracy : 0.77336

'Positive' Class : Yes

CONFUSION MATRIX - rbfDDA – TEST

Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	1067	38
Yes	5	6

Accuracy : 0.9615
95% CI : (0.9484, 0.972)
No Information Rate : 0.9606
P-Value [Acc > NIR] : 0.4787

Kappa : 0.2057
McNemar's Test P-Value : 1.061e-06

Sensitivity : 0.136364
Specificity : 0.995336
Pos Pred Value : 0.545455
Neg Pred Value : 0.965611
Prevalence : 0.039427
Detection Rate : 0.005376
Detection Prevalence : 0.009857
Balanced Accuracy : 0.565850

'Positive' Class : Yes

CONFUSION MATRIX - MXNET - TEST

Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	1071	38
Yes	1	6

Accuracy : 0.9651
95% CI : (0.9525, 0.975)
No Information Rate : 0.9606
P-Value [Acc > NIR] : 0.2483

Kappa : 0.2269
McNemar's Test P-Value : 8.185e-09

Sensitivity : 0.136364
Specificity : 0.999067
Pos Pred Value : 0.857143
Neg Pred Value : 0.965735
Prevalence : 0.039427
Detection Rate : 0.005376
Detection Prevalence : 0.006272
Balanced Accuracy : 0.567715

'Positive' Class : Yes

CONFUSION MATRIX - ENSEMBLE – TEST

Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	1061	31
Yes	11	13

Accuracy : 0.9624
95% CI : (0.9495, 0.9727)
No Information Rate : 0.9606
P-Value [Acc > NIR] : 0.41755

Kappa : 0.3647
McNemar's Test P-Value : 0.00337

Sensitivity : 0.29545
Specificity : 0.98974
Pos Pred Value : 0.54167
Neg Pred Value : 0.97161
Prevalence : 0.03943
Detection Rate : 0.01165
Detection Prevalence : 0.02151
Balanced Accuracy : 0.64260

'Positive' Class : Yes

In the modelling task result above, we found out that the sophisticated ensemble did not do better than the individual model.

MXNET have a higher accuracy rate than the nnet, rbfDDA and the ensemble. However, NNET have a better sensitivity than the 3 other models which means it can identify the percentage of those who have heart attack better. The specificity of the four models are comparable.

To improve the ensemble model, we can have more data to train the models. We can also select a model that seems to capture different aspects of the data and different models that perform best on different subsets of the data.

Having said that, the ensemble we build has a conservative predication of the problem domain therefore is still relevant to us.

6.0 DEVELOPMENT AND LIMITATION

The accuracy results of the four models are quite optimistic. But, we should not see it as a mandate that the models with higher accuracy have higher predictive power. Lower accuracy can have higher predictive power than one with high accuracy. To avoid this accuracy paradox, we should look at the sensitivity and specificity. However, the sensitivity of the four models were not good.

As the objective was prevention instead of diagnosis of heart attack, we can accept a certain degree of lower sensitivity in predicting heart attack as it is not a mission critical task. But we can improve the model sensitively by including more relevant predictors to target variable. (eg. Blood pressure)

7.0 REFERENCES

Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R (2017). Heart Disease and Stroke Statistics—2017 Update: A Report from the American Heart Association. 135: e1–e458. DOI: 10.1161/CIR.0000000000000485.

National Center for Health Statistics. "National Health and Nutrition Examination Survey." Available from <<http://www.cdc.gov/nchs/nhanes>> Retrieved 28 Feb 2018.

Elizabeth Jackson. (2017). Heart Disease and Stroke Statistics: 2017 Update. Retrieved May 10, 2018, from <http://www.acc.org/latest-in-cardiology/ten-points-to-remember/2017/02/09/14/58/heart-disease-and-stroke-statistics-2017>

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), 1225-1231. doi:10.1016/s0895-4356(96)00002-9

APPENDIX A: CODE VALUES AND DESCRIPTION

Variable Name	Variable Description	Code or value	Value Description
SEQN	Sequence Number	Range of values for ID	ID
RIAGENDR	Gender of the participant.	1,2	1-Male, 2-Female
RIDAGEYR	Age in years of the participant at the time of screening. Individuals 80 and over are topcoded at 80 years of age.	0-80	0 to 79 - Range of ages, 80 - 80 and above
RIDAGEYR	Binned Age	s1,s2,s3,s4	s1-younger than 18, s2-18 to 44, s3-45 to 64, s4-65 and older
BPQ020	{Have you/Has SP} ever been told by a doctor or other health professional that {you/s/he} had hypertension, also called high blood pressure?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
BPQ080	{Have you/Has SP} ever been told by a doctor or other health professional that {your/his/her} blood cholesterol level was high?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DIQ010	The next questions are about specific medical conditions. {Other than during pregnancy, {have you/has SP}/{Have you/Has SP}} ever been told by a doctor or health professional that {you have/{he/she/SP} has} diabetes or sugar diabetes?	1,2,3, 7,9	1-Yes, 2-No, 3-Borderline, 7-Don't know, 9-Refused
DIQ050	{Is SP/Are you} now taking insulin	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DBQ700	Next I have some questions about {your/SP?s} eating habits. In general, how healthy is {your/his/her} overall diet? Would you say . . .	1,2,3,4,5,7,9	1-Excellent, 2-Very Good, 3-Good, 4-Fair, 5-Poor, 7-Don't know, 9-Refused
DLQ010	With this next set of questions, we want to learn about people who have physical, mental, or emotional conditions that cause serious difficulties with their daily activities. Though different, these questions may sound similar to ones I asked earlier. {Are you/Is SP} deaf or {do you/does he/does she} have serious difficulty hearing?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DLQ020	{Are you/Is SP} blind or {do you/does he/does she} have serious difficulty seeing even when wearing glasses?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DLQ040	Because of a physical, mental, or emotional condition, {do you/does he/does she} have serious difficulty concentrating, remembering, or making decisions?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DLQ050	{Do you/Does SP} have serious difficulty walking or climbing stairs?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DLQ060	{Do you/Does SP} have difficulty dressing or bathing?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
HEQ010	Has a doctor or other health professional ever told {you/SP} that {you have/s/he/SP has} Hepatitis B? (Hepatitis is a form of liver disease. Hepatitis B is an infection of the liver from the Hepatitis B virus (HBV).)	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
HEQ030	Has a doctor or other health professional ever told {you/SP} that {you have/s/he/SP	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused

Variable Name	Variable Description	Code or value	Value Description
	has} Hepatitis C? (Hepatitis is a form of liver disease. Hepatitis C is an infection of the liver from the Hepatitis C virus (HCV).)		
MCQ010	The following questions are about different medical conditions. Has a doctor or other health professional ever told {you/SP} that {you have/s/he/SP has} asthma (az-ma)?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ053	During the past 3 months, {have you/has SP} been on treatment for anemia (a-nee-me-a), sometimes called "tired blood" or "low blood"? [Include diet, iron pills, iron shots, transfusions as treatment.]	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ080	Has a doctor or other health professional ever told {you/SP} that {you were/s/he/SP was} overweight?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ082	Has a doctor or other health professional ever told {you/SP} that {you have/s/he/SP has} celiac (sele-ak) disease, also called or sprue (sproo)?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ086	{Are you/is SP} on a gluten-free diet?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160A	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had arthritis (ar-thry-tis)?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160N	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had gout?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160B	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had congestive heart failure?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160C	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had coronary (kor-o-nare-ee) heart disease?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160D	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had angina (an-gi-na), also called angina pectoris?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160E	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had a heart attack (also called myocardial infarction (my-o-car-dee-al in-fark-shun))?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160F	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had a stroke?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160G	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had emphysema (emph-phi-see-ma)?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160M	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had another thyroid (thigh-roid) problem?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160K	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had chronic bronchitis?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160L	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had any kind of liver condition?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ160O	Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . .had COPD?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ220	{Have you/Has SP} ever been told by a doctor or other health professional that {you/s/he}	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused

Variable Name	Variable Description	Code or value	Value Description
	had cancer or a malignancy (ma-lig-nan-see) of any kind?		
MCQ365A	To lower {your/SP's} risk for certain diseases, during the past 12 months {have you/has s/he} ever been told by a doctor or health professional to: control {your/his/her} weight or lose weight?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ365B	To lower {your/SP's} risk for certain diseases, during the past 12 months {have you/has s/he} ever been told by a doctor or health professional to: increase {your/his/her} physical activity or exercise?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ365C	To lower {your/SP's} risk for certain diseases, during the past 12 months {have you/has s/he} ever been told by a doctor or health professional to: reduce the amount of sodium or salt in {your/his/her} diet?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ365D	To lower {your/SP's} risk for certain diseases, during the past 12 months {have you/has s/he} ever been told by a doctor or health professional to: reduce the amount of fat or calories in {your/his/her} diet?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ370A	To lower {your/his/her} risk for certain diseases, {are you/is s/he} now doing any of the following: controlling {your/his/her} weight or losing weight?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ370B	To lower {your/his/her} risk for certain diseases, {are you/is s/he} now doing any of the following: increasing {your/his/her} physical activity or exercise?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ370C	To lower {your/his/her} risk for certain diseases, {are you/is s/he} now doing any of the following: reducing the amount of sodium or salt in {your/his/her} diet?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ370D	To lower {your/his/her} risk for certain diseases, {are you/is s/he} now doing any of the following: reducing the amount of fat or calories in {your/his/her} diet?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
PFQ090	{Do you/Does SP} now have any health problem that requires {you/him/her} to use special equipment, such as a cane, a wheelchair, a special bed, or a special telephone?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
PAQ605	Next I am going to ask you about the time {you spend/SP spends} doing different types of physical activity in a typical week. Think first about the time {you spend/he spends/she spends} doing work. Think of work as the things that {you have/he has/she has} to do such as paid or unpaid work, household chores, and yard work. Does {your/SP's} work involve vigorous-intensity activity that causes large increases in breathing or heart rate like carrying or lifting heavy loads, digging or construction work for at least 10 minutes continuously?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
PAQ620	Does {your/SP's} work involve moderate-intensity activity that causes small increases in breathing or heart rate such as brisk walking or carrying light loads for at least 10 minutes continuously?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused

Variable Name	Variable Description	Code or value	Value Description
PAQ635	The next questions exclude the physical activity at work that you have already mentioned. Now I would like to ask you about the usual way {you travel/SP travels} to and from places. For example to school, for shopping, to work. In a typical week {do you/does SP} walk or use a bicycle for at least 10 minutes continuously to get to and from places?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
PAQ650	The next questions exclude the work and transport activities that you have already mentioned. Now I would like to ask you about sports, fitness and recreational activities. In a typical week {do you/does SP} do any vigorous-intensity sports, fitness, or recreational activities that cause large increases in breathing or heart rate like running or basketball for at least 10 minutes continuously?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
PAQ665	In a typical week {do you/does SP} do any moderate-intensity sports, fitness, or recreational activities that cause a small increase in breathing or heart rate such as brisk walking, bicycling, swimming, or volleyball for at least 10 minutes continuously?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
SLQ050	{Have you/Has SP} ever told a doctor or other health professional that {you have/s/he has} trouble sleeping?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
SLQ060	{Have you/Has SP} ever been told by a doctor or other health professional that {you have/s/he has} a sleep disorder?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
SMQ020	These next questions are about cigarette smoking and other tobacco use. {Have you/Has SP} smoked at least 100 cigarettes in {your/his/her} entire life?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DBD900		1-21, 0, 5555, 7777, 9999	1 to 21 - Range of values, 0-None, 5555, More than 21, 7777-Refused, 9999-Don't know
PAQ710	Now I will ask you first about TV watching and then about computer use. Over the past 30 days, on average how many hours per day did {you/SP} sit and watch TV or videos? Would you say . . .	0,1,2,3,4,5,8,77,99	0-Less than 1 hour, 1-1hour, 2-2hour,3-3hour,4-4hour, 5-hours or more, 8-don't watch, 77-refused, 99-don't know
PAQ715	Over the past 30 days, on average how many hours per day did {you/SP} use a computer or play computer games outside of school? Include Playstation, Nintendo DS, or other portable video games Would you say . . .	0,1,2,3,4,5,8,77,99	0-Less than 1 hour, 1-1hour, 2-2hour,3-3hour,4-4hour, 5-hours or more, 8-don't watch, 77-refused, 99-don't know
SLD010H	The next set of questions is about your sleeping habits. How much sleep {do you/does SP} usually get at night on weekdays or workdays?	2-11, 12, 77, 99	2 to 11-Range of values, 12-12hours or more, 77-refused, 99-don't know
DR1TKCAL	Energy (kcal)	Numeric values	Range of values
DR1TPROT	Protein (gm)	Numeric values	Range of values
DR1TCARB	Carbohydrate (gm)	Numeric values	Range of values
DR1TSUGR	Total sugars (gm)	Numeric values	Range of values
DR1TFIBE	Dietary fiber (gm)	Numeric values	Range of values

Variable Name	Variable Description	Code or value	Value Description
DR1TTFAT	Total fat (gm)	Numeric values	Range of values
DR1TSFAT	Total saturated fatty acids (gm)	Numeric values	Range of values
DR1TMFAT	Total monounsaturated fatty acids (gm)	Numeric values	Range of values
DR1TPFAT	Total polyunsaturated fatty acids (gm)	Numeric values	Range of values
DR1TCHOL	Cholesterol (mg)	Numeric values	Range of values
BMXWT	Weight (kg)	Numeric values	Range of values
BMXHT	Standing Height (cm)	Numeric values	Range of values
BMXWAIST	Waist Circumference (cm)	Numeric values	Range of values
BMXBMI	Body Mass Index (kg/m**2)	Numeric values	Range of values
BMI_Cal_Cat	BMI Category - Based on CDC	Underweight, Normal, Overweight, Obese	BMI below 18.5-Underweight; 18-5 to 24.9-Normal; 25.0 to 29.9-Overweight; 30.0 and above-Obese
ALQ101	The next questions are about drinking alcoholic beverages. Included are liquor (such as whiskey or gin), beer, wine, wine coolers, and any other type of alcoholic beverage. In any one year, {have you/has SP} had at least 12 drinks of any type of alcoholic beverage? By a drink, I mean a 12 oz. beer, a 5 oz. glass of wine, or one and half ounces of liquor.	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
HSD010	Next I have some general questions about {your/SP's} health. Would you say {your/SP's} health in general is . . .	1,2,3,4,5,7,9	1-Excellent, 2-Very Good, 3-Good, 4-Fair, 5-Poor, 7-Don't know, 9-Refused
DIQ170	{Have you/Has SP} ever been told by a doctor or other health professional that {you have/s/he has} health conditions or a medical or family history that increases {your/his/her} risk for diabetes?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DPQ010	Over the last 2 weeks, how often have you been bothered by the following problems: little interest or pleasure in doing things? Would you say...	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ020	[Over the last 2 weeks, how often have you been bothered by the following problems:] feeling down, depressed, or hopeless?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ030	[Over the last 2 weeks, how often have you been bothered by the following problems:] trouble falling or staying asleep, or sleeping too much?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ040	[Over the last 2 weeks, how often have you been bothered by the following problems:] feeling tired or having little energy?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ050	[Over the last 2 weeks, how often have you been bothered by the following problems:] poor appetite or overeating?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ060	[Over the last 2 weeks, how often have you been bothered by the following problems:] feeling bad about yourself - or that you are a failure or have let yourself or your family down?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ070	[Over the last 2 weeks, how often have you been bothered by the following problems:]	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-

Variable Name	Variable Description	Code or value	Value Description
	trouble concentrating on things, such as reading the newspaper or watching TV?		Nearly every day, 7-Refused, 9-Don't know
DPQ080	[Over the last 2 weeks, how often have you been bothered by the following problems:] moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
DPQ090	Over the last 2 weeks, how often have you been bothered by the following problems: Thoughts that you would be better off dead or of hurting yourself in some way?	0,1,2,3,7,9	0-Not at all, 1-Several days, 2-More than half the days, 3-Nearly every day, 7-Refused, 9-Don't know
MCQ300A	Including living and deceased, were any of {SP's/your} close biological that is, blood relatives including father, mother, sisters or brothers, ever told by a health professional that they had a heart attack or angina (an-gi-na) before the age of 50?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ300B	Including living and deceased, were any of {SP's/your} close biological that is, blood relatives including father, mother, sisters or brothers, ever told by a health professional that they had asthma (az-ma)?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
MCQ300C	Including living and deceased, were any of {SP's/your} close biological that is, blood relatives including father, mother, sisters or brothers, ever told by a health professional that they had diabetes?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
RXQ510	Doctors and other health care providers sometimes recommend that {you take/SP takes} a low-dose aspirin each day to prevent heart attacks, strokes, or cancer. {Have you/Has SP} ever been told to do this?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused
DRQSDIET	Are you currently on any kind of diet, either to lose weight or for some other health-related reason?	1,2,7,9	1-Yes, 2-No, 7-Don't know, 9-Refused