ISSN: 2689-1204

Research Article

Journal of Applied Material Science & Engineering Research

Applying Multiple Regression Analyses to Compare the Regression Predicted and Originally Calculated CVD/Stroke Risk Probabilities Using Medical Condition and Lifestyle Detail Scores as Inputs over a 10-year Period for a type 2 Diabetes Patient Based on GH-Method: Math-Physical Medicine (No. 546)

Gerald C Hsu

EclaireMD Foundation, USA

***Corresponding author**
Gerald C Hsu, EclaireMD Foundation, USA

Submitted: 01 Dec 2021; **Accepted:** 08 Dec 2021; **Published:** 17 Dec 2021

Citation: *Gerald C Hsu (2021) Applying Multiple Regression Analyses to Compare the Regression Predicted and Originally Calculated CVD/Stroke Risk Probabilities Using Medical Condition and Lifestyle Detail Scores as Inputs over a 10-year Period for a type 2 Diabetes Patient Based on GH-Method: Math-Physical Medicine (No. 546). J App Mat Sci & Engg Res, 5(4), 01-06.*

Abstract

In the author's previous medical research reports, he mainly applied physics theories, engineering models, mathematical equations, computer big data analytics and artificial intelligence (AI) techniques, as well as some statistical approaches to explore and interpret various biophysical phenomena. However, the majority of medical research papers he has read thus far are primarily based on statistics. As a result, in this article, he selects some basic statistical tools, such as correlation, variance, p-values, and multiple regression analyses, to study the predicted CVD/Stroke risk probability as the output (dependent variable) by using his medical condition and lifestyle detail scores as inputs (independent variables).

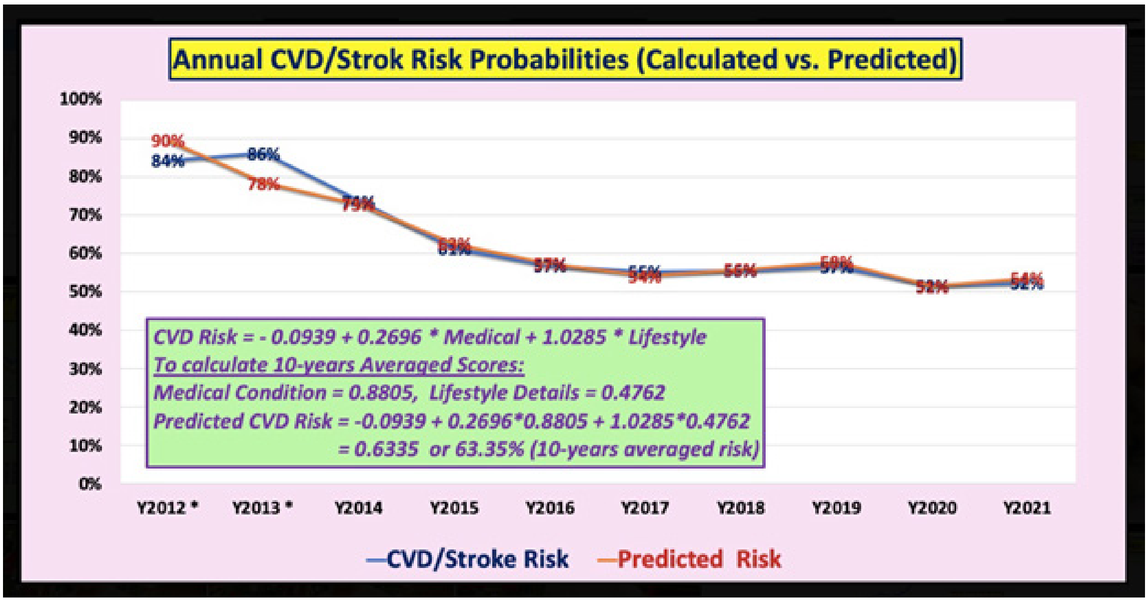
Since 1/1/2012, the author has been collecting various data related to his health (~3 million data) which includes 4 categories of medical conditions, obesity, diabetes, hypertension, and hyperlipidemia (m1 through m4), along with 6 categories of lifestyle details, including exercise, water intake, sleep, stress, food, and daily life routines (m6 through m10). However, due to the limitation of his learned knowledge in earlier years, the data from 2012-2013 is incomplete; therefore, the scores in this study for the initial period of 2012-2013 are his best-guessed data based on an incomplete dataset.

Previously, he has researched and published a few articles regarding the risks of having CVD/Stroke based on his metabolism index (MI). In this paper, he will compare his previously calculated CVD risks based on the MI method versus the newly calculated CVD risks based on multiple regression analysis methods.

In this study, he will not repeat the detailed introduction of the regression analysis in the Method section because it is available in many statistics textbook. It should be noted that in regression analysis, the correlation coefficient R should be > 0.5 or 50% to indicate a strong inter-connectivity and the p-value should be <0.05 to be considered as statistically significant.

In summary, there are three specific conclusions worth mentioning:

1. *The multiple regression based predicted CVD Risk (orange curve) and previously calculated CVD Risk (blue curve) based on MI have an extremely strong correlation (R=97%) and high variance (R^2=94%). In addition, it has a 100% prediction accuracy rate. These findings have proven the usefulness and accuracy of the predicted dependent variable, CVD Risk, by using multiple regression analysis results of 2 independent variables as inputs, medical condition and lifestyle detail scores.*
2. *Furthermore, he has analyzed the connectivity between CVD Risk vs. medical conditions (R= 85%, R^2=73%) and CVD Risk vs. lifestyle details (R= 96%, R^2=93%). It is obvious that, in his case, lifestyle management contributes more than existing medical conditions to his risk probability of developing CVD/Stroke in the near future. This conclusion is based on those observed higher correlation and lower p-value associated width lifestyle. This is logical to him because bad lifestyle habits result in chronic diseases which further develop CVD/Stroke complication. Lifestyle is the root cause of all chronic diseases and their complications.*
3. *Using the CVD/Stroke prediction equation based on multiple regression analyses, it has obtained a 63.35% average CVD Risk over 10 years which is the same as the previously calculated CVD Risk of 63.35% based on the MI method.*



Introduction

In the author's previous medical research reports, he mainly applied physics theories, engineering models, mathematical equations, computer big data analytics and artificial intelligence (AI) techniques, as well as some statistical approaches to explore and interpret various biophysical phenomena. However, the majority of medical research papers he has read thus far are primarily based on statistics. As a result, in this article, he selects some basic statistical tools, such as correlation, variance, p-values, and multiple regression analyses, to study the predicted CVD/Stroke risk probability as the output (dependent variable) by using his medical condition and lifestyle detail scores as inputs (independent variables).

Since 1/1/2012, the author has been collecting various data related to his health (~3 million data) which includes 4 categories of medical conditions, obesity, diabetes, hypertension, and hyperlipidemia (m1 through m4), along with 6 categories of lifestyle details, including exercise, water intake, sleep, stress, food, and daily life routines (m6 through m10). However, due to the limitation of his learned knowledge in earlier years, the data from 2012-2013 is incomplete; therefore, the scores in this study for the initial period of 2012-2013 are his best-guessed data based on an incomplete dataset.

Previously, he has researched and published a few articles regarding the risks of having CVD/Stroke based on his metabolism index (MI). *In this paper, he will compare his previously calculated CVD risks based on the MI method versus the newly calculated CVD risks based on multiple regression analysis methods.*

In this study, he will not repeat the detailed introduction of the regression analysis in the Method section because it is available in many statistics textbook. It should be noted that in regression analysis, the correlation coefficient R should be > 0.5 or 50% to indicate a strong inter-connectivity and the p-value should be <0.05 to be considered as statistically significant.

Methods

MPM Background

To learn more about the author's developed GH-Method: math-physical medicine (MPM) methodology, readers can select the following three papers from his ~500 published medical papers.

The first paper, No. 386 describes his MPM methodology in a general conceptual format. The second paper, No. 387 outlines the history of his personalized diabetes research, various application tools, and the differences between biochemical medicine (BCM) approach versus the MPM approach. The third paper, No. 397 depicts a general flow diagram containing ~10 key MPM research methods and different tools.

In particular, paper No. 453 illustrates his GH-Method: math-physical medicine in great details, "Using Topology concept of mathematics and Finite Element method of engineering to develop a mathematical model of Metabolism in medicine in order to control various chronic diseases and their complications via overall health conditions improvement".

The Author's Case of Diabetes and Complications

The author has been a severe type 2 diabetes (T2D) patient since 1996 and weighed 220 lbs. (100 kg, BMI 32.5) at that time. By 2010, he still weighed 198 lbs. (BMI 29.2) with an average daily glucose of 250 mg/dL (HbA1C of 10%). During that year, his triglycerides reached to 1161 (diabetic retinopathy or DR) and albumin-creatinine ratio (ACR) at 116 (chronic kidney disease or CKD). He also suffered five cardiac episodes within a decade. In 2010, three independent physicians warned him regarding his needs of kidney dialysis treatment and future high risk of dying from severe diabetic complications. Other than cerebrovascular disease (stroke), he has suffered most known diabetic complications, including both macro-vascular and micro-vascular complications.

In 2010, he decided to launch his self-study on endocrinology, diabetes, and food nutrition in order to save his own life. During 2015 and 2016, he developed four prediction models related to diabetes conditions: weight, postprandial plasma glucose (PPG), fasting plasma glucose (FPG), and A1C. As a result, from using his developed mathematical metabolism index (MI) model in 2014 and the four prediction tools, by end of 2016, his weight was reduced from 220 lbs. (100 kg, BMI 32.5) to 176 lbs. (89 kg, BMI 26.0), waistline from 44 inches (112 cm, nonalcoholic fatty liver disease /NAFLD) to 33 inches (84 cm), average finger glucose reading from 250 mg/dL to 120 mg/dL, and lab-tested A1C from 10% to ~6.5%. One of his major accomplishments is that he no longer takes any diabetes medications since 12/8/2015.

In 2017, he has achieved excellent results on all fronts, especially his glucose control. However, during the pre-COVID period of 2018 and 2019, he traveled to approximately 50+ international cities to attend 65+ medical conferences and made ~120 oral presentations. This hectic schedule inflicted damage to his diabetes control, through dinning out frequently, post-meal exercise disruption, jet lag, and along with the overall metabolism impact due to his irregular life patterns through a busy travel schedule; therefore, his glucose control and overall metabolism state were somewhat affected during this two-year heavy travel period.

During 2020 and 2021 with a strict COVID-19 quarantine lifestyle, not only has he written and published ~400 medical papers in 100+ journals, but he has also reached his best health conditions for the past 26 years. By the beginning of 2021, his weight was further reduced to 165 lbs. (BMI 24.4) along with a 6.1% A1C value (daily average glucose at 105 mg/dL), without having any medication interventions or insulin injections. These good results are due to his non-traveling, low-stress, and regular daily life routines. Due to the knowledge of chronic diseases, practical lifestyle management experiences, and his developed various high-tech tools, they contributed to his excellent health status since 1/19/2020, which is the start date of being self-quarantine.

On 5/5/2018, he applied a CGM sensor device on his upper arm and checks glucose measurements every 5 minutes for a total of ~288 times each day. He has maintained the same measurement pattern to present day. In his research work, he uses the CGM sensor glucose at time-interval of 15 minutes (96 data per day). Incidentally, the difference of average sensor glucoses between 5-minute intervals and 15-minute intervals is only 0.4% (average glucose of 114.81 mg/dL for 5-minutes and average glucose of 114.35 mg/dL for 15-minutes with a correlation of 93% between these two sensor glucose curves) during the period from 2/19/20 to 8/13/21.

Therefore, over the past 11 years, he could study and analyze the collected ~3 million data regarding his health status, medical conditions, and lifestyle details. He applies his knowledge, models, and tools from mathematics, physics, engineering, and computer science to conduct his medical research work. His medical research work is based on the aims of achieving both "high precision" with "quantitative proof" in the medical findings.

The following timetable provides a rough sketch of the emphasis of his medical research during each stage:

- 2000-2013: Self-study diabetes and food nutrition, developing a data collection and analysis software.
- 2014: Develop a mathematical model of metabolism, using engineering modeling and advanced mathematics.
- 2015: Weight & FPG prediction models, using neuroscience.
- 2016: PPG & HbA1C prediction models, utilizing optical physics, AI, and neuroscience.
- 2017: Complications due to macro-vascular research such as cardiovascular disease (CVD), coronary heart disease (CHD) and stroke, using pattern analysis and segmentation analysis.
- 2018: Complications due to micro-vascular research such as CKD, bladder, foot, and eye issues such as DR.
- 2019: CGM big data analysis, using wave theory, energy theory, frequency domain analysis, quantum mechanics, and AI.
- 2020: Cancer, dementia, longevity, geriatrics, DR, hypothyroidism, diabetic foot, diabetic fungal infection, linkage between metabolism and immunity, and learning about certain infectious diseases such as COVID-19.
- 2021: Applications of LEGT and perturbation theory from quantum mechanics on medical research subjects, such as chronic diseases and their complications, cancer, and dementia. Using metabolism and immunity.it's as the base, he expands his research into cancers, dementia, and COVID-19. In addition, he has also developed a few useful analysis methods and tools for his medical research work.

To date, he has collected nearly 3 million data regarding his medical conditions and lifestyle details. In addition, he has written 536 medical papers and published 500+ articles in 100+ various medical journals, including 7 special editions with selected 20-25 papers for each edition. Moreover, he has given ~120 presentations at ~65 international medical conferences. He has continuously dedicated time and effort on medical research work to share his findings and knowledge with patients worldwide.

His Risk Probability Model of CVD

In this paper, the author described how to apply his engineering science background, including mathematics, physics, and computer science to conduct his medical research about risk probability of having various complications of chronic diseases, such as stroke, CVD, CHD, CKD, DK, dementia, and cancer. Here, he specifically describes the risk model of having a CVD or stroke.

He reviewed his collected data of 6 years from 10/12/2016 through 10/11/2021, where he focused on his 4 chronic disease's medical conditions, including obesity, diabetes, hypertension, and hyperlipidemia, along with the medical conditions plus his 6 lifestyle details.

As a part of his medical research, he applied the acquired mechanical and structural engineering knowledge to develop several biomedical scenarios to research the chronic diseases, obesity, diabetes, hypertension, and hyperlipidemia, along with their induced various complications. One of these complications is CVD/stroke.

The engineering analogy of deaths caused by disease and human expected lifespan can be explained simply by using an example of a new machine or a new bridge. If we develop a monitoring system to continuously measure, record, and analyze the external forces, material strength, and material damages for a machine or bridge, as well as the relationship between *force/stress* (*similar to causes* of a disease such as lifestyle details) and *deformation/strain* (*similar to symptoms* of a disease such as medical conditions), we can then have a clear idea how severe the damages are and to determine the useful life or expected lifespan of the machine or bridge.

The author self-studied chronic diseases, metabolism, and food nutrition for 4-years from 2010 to 2013. He started his medical research work by building a mathematical metabolism model in 2014. He named his research methodology as the "GH-method: math-physical medicine (MPM approach)". Over the past 11 years of his medical research work, he has learned that the most important factor is knowing how to apply physics principles and engineering modeling techniques to various biomedical problems. This is different from simply inserting your biomedical data into some existing mathematical equations extended from physical theories and engineering models.

The reason for doing this is that the original mathematical equations associated with the original inventors' theories or models usually come with their inherited boundary conditions. These conditions may or may not fit perfectly with the biomedical situations directly; therefore, you must understand the scope and applicability of these physical theories and engineering models first, and then find a suitable way to apply them. In other words, by learning other people's wisdom first and then find a way to apply their wisdom to your own biomedical problem is the most practical way to solve these different biomedical problems.

The author's simple numerical calculation of risk probability is based on his knowledge and applications of physics law/concept and engineering modeling technique, big data analytics, and his developed mathematical metabolism model. It has depicted a viable way to extend lowering the risk probability of having various complications through an effective metabolic condition improvement and lifestyle maintenance program. This practical method has already been applied and proven effectively in controlling his type 2 diabetes (T2D) and its various complications without taking medications for the past 6 years. In his disease risk model, it includes basic conditions, such as genetic and life-time bad habits. When dealing with medical conditions, for example, the artery blockage situations are mainly related to glucose and blood lipids, while artery fracture situations are primarily related to glucose and blood pressure. Furthermore, his 6 lifestyle details, food & diet, water intake, exercise, sleep, stress, and daily life routines also play vital roles in the determination of CVD/Stroke risks.

Results

Figure 1 displays a summarized data table of the multiple regression analyses of his CVD/Stroke risk versus the annual medical conditions and annual lifestyle detail scores over a 10-year period from 2012 to 2021. There are 10 observations (years) with the significance F value of 0.00006; therefore, the results are statistically significant.

The 5 key data are listed as follows:

Correlation (R) = 97%
Variance (R^2) = 94%
MI averaged risk = 63%
Regression Predicted risk = 63%
Prediction accuracy = 100%

11/10/21	Y	X1	X2			
CVD vs. M1-M4 & M5-M10	CVD/Stroke Risk	Medical Conditions	Lifestyle	CVD vs. Metabolism	CVD/Stroke Risk	Predicted Risk
Y2012 *	0.8428	0.9999	0.7000	Y2012 *	84%	90%
Y2013 *	0.8615	0.9659	0.6000	Y2013 *	86%	78%
Y2014	0.7364	0.9831	0.5407	Y2014	74%	73%
Y2015	0.6123	0.9471	0.451	Y2015	61%	63%
Y2016	0.5669	0.8188	0.4341	Y2016	57%	57%
Y2017	0.5533	0.7809	0.4132	Y2017	55%	54%
Y2018	0.5544	0.8264	0.4174	Y2018	55%	56%
Y2019	0.5653	0.8685	0.4254	Y2019	57%	58%
Y2020	0.5162	0.7581	0.3921	Y2020	52%	51%
Y2021	0.5234	0.8562	0.3876	Y2021	52%	54%
Average	0.6333	0.8805	0.4762	Average	63%	63%
Correlation	100%	85%	96%	Correlation	97%	
Variance	100%	73%	93%	Variance	94%	
Slope		0.0023	1.5915			
Y-Intercept	-0.1102					

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9684
R Square	0.9378
Adjusted R Square	0.9200
Standard Error	0.0371
Observations	10

ANOVA	df	SS	MS	F	Significance F
Regression	2	0.144872457	0.072436229	52.75100392	0.00006
Residual	7	0.009612208	0.001373173		
Total	9	0.154484665			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-0.0939	0.148768195	-0.63090896	0.5481	-0.44564007	0.257921695	-0.44564007	0.257921695
Medical Conditions	0.2696	0.249354533	1.081161123	0.3155	-0.320037349	0.859222203	-0.320037349	0.859222203
Lifestyle	1.0285	0.210225201	4.892527146	0.0018	0.531428894	1.525636112	0.531428894	1.525636112

RESIDUAL OUTPUT

Observation	Predicted CVD/Stroke Risk	Residuals
1	90%	-0.052879033
2	78%	0.07784036
3	73%	0.009095348
4	63%	-0.013039959
5	57%	-0.006469051
6	54%	0.011644831
7	56%	-0.003841461
8	58%	-0.012519562
9	51%	0.002393574
10	54%	-0.012225047

Figure 1: Data table of multiple regression analysis results of CVD risk resulted from both medical conditions and lifestyle details (2012-2021)

Figure 2 illustrates the comparison of his multiple regression analysis CVD risk (orange curve) versus his previously calculated CVD risk (blue curve) based on the metabolism index (MI) model. It is very clear that his regression predicted CVD risks are almost identical to the metabolism calculated CVD risks, except for the initial period of 2012-2013 due to the data integrity issue.

As a reference, here is the multiple regression generated annual CVD/Stroke risk equation :

Y (predicted CVD risk)
= - 0.0939 + 0.2696* X1 (medical condition score) + 1.0285*X2 (lifestyle detail score)

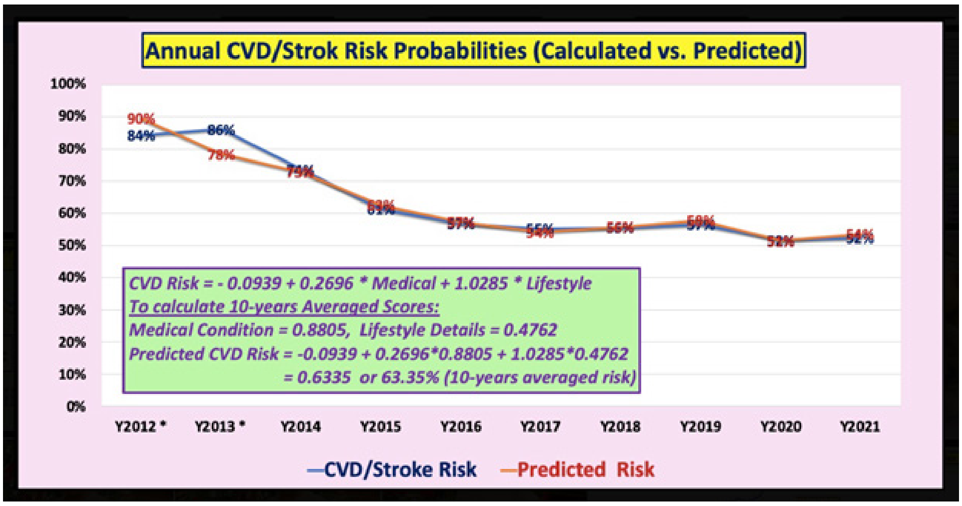


Figure 2: Metabolism Index based CVD/Stroke risk versus Multiple Regression based Predicted CVD/Stroke risk using medical condition scores and lifestyle detail scores as inputs (2012-2021)

Figure 3 shows two diagrams of multiple regressions based predicted CVD/Stroke risk versus medical conditions and lifestyle detail score as inputs (2012-2021). The CVD risk is more dependent on lifestyle (higher R and R^2) than medical conditions (lower correlation and variance).

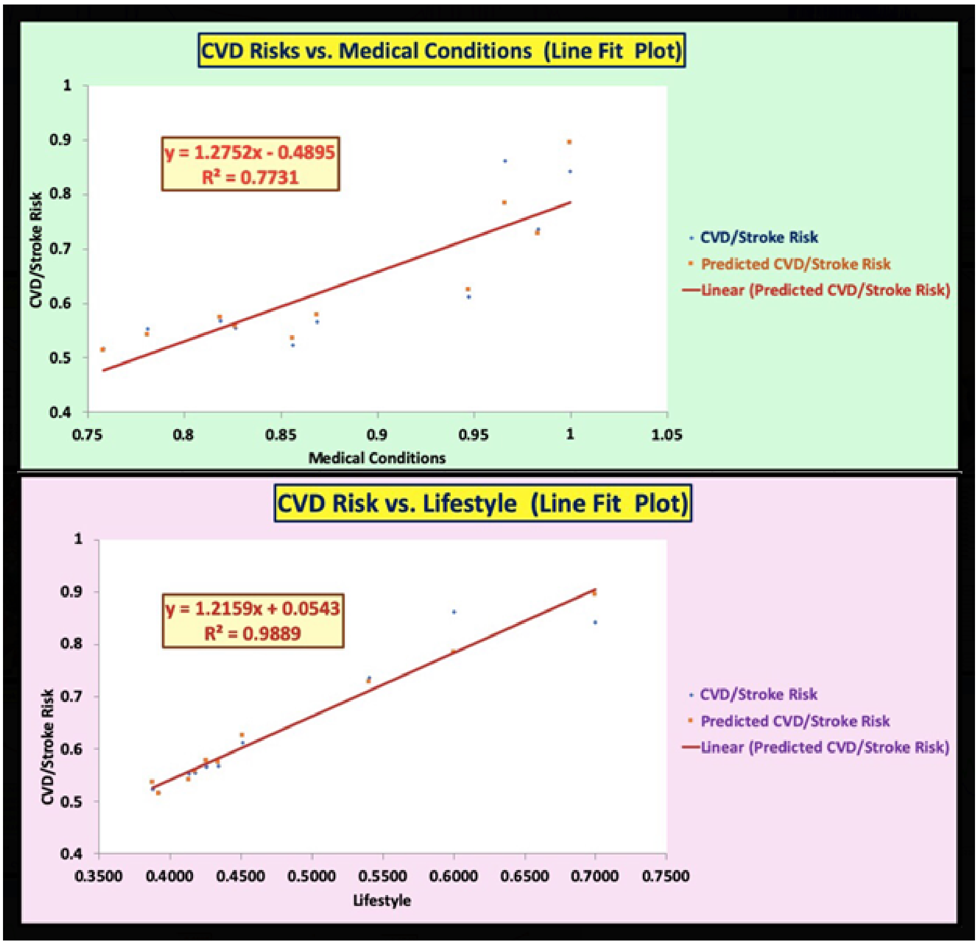


Figure 3: Multiple Regression based Predicted CVD/Stroke risk versus both medical conditions score and lifestyle detail scores as inputs (2012-2021)

Conclusions

In summary, there are three specific conclusions worth mentioning:

1. The multiple regression based predicted CVD Risk (orange curve) and previously calculated CVD Risk (blue curve) based on MI have an extremely strong correlation (R=97%) and high variance (R^2=94%). In addition, it has a 100% prediction accuracy rate. ***These findings have proven the usefulness and accuracy of the predicted dependent variable, CVD Risk, by using multiple regression analysis results of 2 independent variables as inputs, medical condition and lifestyle detail scores.***
2. Furthermore, he has analyzed the connectivity between CVD Risk vs. medical conditions (R= 85%, R^2=73%) and CVD Risk vs. lifestyle details (R= 96%, R^2=93%). It is obvious that, in his case, lifestyle management contributes more than existing medical conditions to his risk probability of developing CVD/Stroke in the near future. This conclusion is based on those observed higher correlation and lower p-value associated width lifestyle. This is logical to him because ***bad lifestyle habits result in chronic diseases which further develop CVD/Stroke complication. Lifestyle is the root cause of all chronic diseases and their complications.***
3. Using the CVD/Stroke prediction equation based on multiple regression analyses, it has obtained a 63.35% average CVD Risk over 10 years which is the same as the previously calculated CVD Risk of 63.35% based on the MI method.

References

For editing purposes, majority of the references in this paper, which are self-references, have been removed for this article. Only references from other authors' published sources remain. The bibliography of the author’s original self-references can be viewed at www.eclairemd.com.

Readers may use this article as long as the work is properly cited, and their use is educational and not for profit, and the author’s original work is not altered.