ISSN: 2689-1204

Research Article

Journal of Applied Material Science & Engineering Research

Applying Linear Regression Analysis Model to Compare the Outputs of 3 Regression Predicted CVD/Stroke Risk Probabilities Using 3 Different Inputs Which are the Calculated Sensor HbA1C value over 14-Months, Combined Medical Condition Score Months, and Calculated Finger HbA1C over an 8-Year Period from the Collected Data of a type 2 Diabetes Patient Based on GH-Method: Math-Physical Medicine (No. 557)

Gerald C Hsu

EclaireMD Foundation, USA

*Corresponding author
Gerald C Hsu, EclaireMD Foundation, USA

Submitted: 01 Dec 2021; **Accepted:** 08 Dec 2021; **Published:** 28 Dec 2021

Note: *Readers who want to get a quick overview can read the abstract, results and graphs sections.*

Citation: *Gerald C Hsu (2021) Applying Linear Regression Analysis Model to Compare the Outputs of 3 Regression Predicted CVD/Stroke Risk Probabilities Using 3 Different Inputs Which are the Calculated Sensor HbA1C value over 14-Months, Combined Medical Condition Score Months, and Calculated Finger HbA1C over an 8-Year Period from the Collected Data of a type 2 Diabetes Patient Based on GH-Method: Math-Physical Medicine (No. 557). J App Mat Sci & Engg Res, 5(4), 01-07.*

Abstract

Since 1/1/2012, the author has been collecting various biomedical and lifestyle of ~3 million data related to his health conditions. This includes the medical categories for 4 chronic diseases of 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).

In early 2018, he studied, researched, and published many articles regarding the risks of having CVD/Stroke based on his developed metabolism index (MI) model. In this paper, he will compare the calculated CVD risks based on the MI model through his developed GH-method: math-physical medicine methodology against the recently calculated 3 CVD risk probabilities based on a traditional statistical regression model but using 3 different input datasets.

In this article, 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.

The main purpose is to distinguish the degree of influences from "diabetes alone" and "combined 4 chronic diseases" on his risk probability of having an episode of CVD or Stroke during the same time period. Within the category of diabetes alone or HbA1C, he also differentiates the results based on the different time periods of the collected data: one 14-month period along with an 8-year period.

In conclusion, his risk of having a CVD/Stroke is highly connected to the combined medical conditions, weight, glucose, blood lipids, blood pressure and heart rate, during the recent 14-month period. However, his risk of having a CVD/Stroke is not strongly related to the main diabetes measuring biomarker of HbA1C during the same period.

Of course, ***any analysis work using various statistical tools must pay attention on the section of the dataset and time period. Within two different time windows, the data distribution pattern and the data variability may differ according to the selected time-window; therefore, the analysis results and conclusions can vary.***

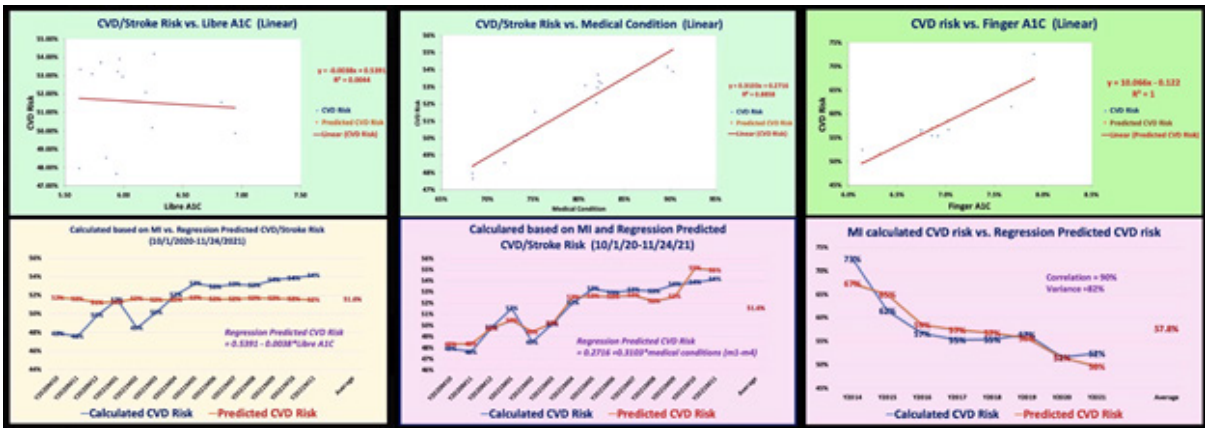
The above observed two different conclusions based on the same 14-month period is due to the fact that ***his glucoses have been under stringent control (with an average HbA1C of 6.1%).*** *Therefore, the variance during the 14-months using A1C as the input is a miniscule 0.4% compared to the variance using medical condition as input is an exceptionally high 89%.*

If we examine the space-domain diagrams closely, using HbA1C as the input, the data results in the scattered map are spread out all over while a straight trend-line has a high difficulty to represent or simulate the majority of his CVD risk data. That is why its variance (R^2) is a mere 0.4%; therefore, the predicted CVD curve using A1C as the input is almost a horizontal line (similar to his A1C curve) and completely out-of-synch with the MI-model calculated CVD risk curve.

On the contrary, by using the medical condition as the input, the scattered results are located within a narrow data bend from the lower left corner to the upper right corner, whereas the straight trend-line represents 89% of the total CVD risk data. As a result, the predicted CVD curve using the medical condition as input is almost identical with the MI-model calculated CVD risk curve in the time-domain chart.

Furthermore, if he uses his finger A1C from the past 8 years (2014 to 2021), the scattered results are also located within a narrow data bend from the lower left corner to the upper right corner, while the straight trend-line represents 82% of the total CVD risk data. This conclusion is a result of the A1C curve trend which fluctuates and matches his CVD risk waveform from the past 8 years.

The findings from this article have demonstrated the importance of selection of both data and time-window which directly influence the final results and conclusions, if we uses statistical methods as our biomedical research tools.



Introduction

Since 1/1/2012, the author has been collecting various biomedical and lifestyle of ~3 million data related to his health conditions. This includes the medical categories for 4 chronic diseases of 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).

In early 2018, he studied, researched, and published many articles regarding the risks of having CVD/Stroke based on his developed metabolism index (MI) model. *In this paper, he will compare the calculated CVD risks based on the MI model through his developed GH-method: math-physical medicine methodology against the recently calculated 3 CVD risk probabilities based on a traditional statistical regression model but using 3 different input datasets.*

In this article, 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.

The main purpose is to distinguish the degree of influences from "diabetes alone" and "combined 4 chronic diseases" on his risk probability of having an episode of CVD or Stroke during the same time period. Within the category of diabetes alone or HbA1C, he also differentiates the results based on the different time periods of the collected data: one 14-month period along with an 8-year period.

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 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 linear elastic glucose theory (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 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 f*orce/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.

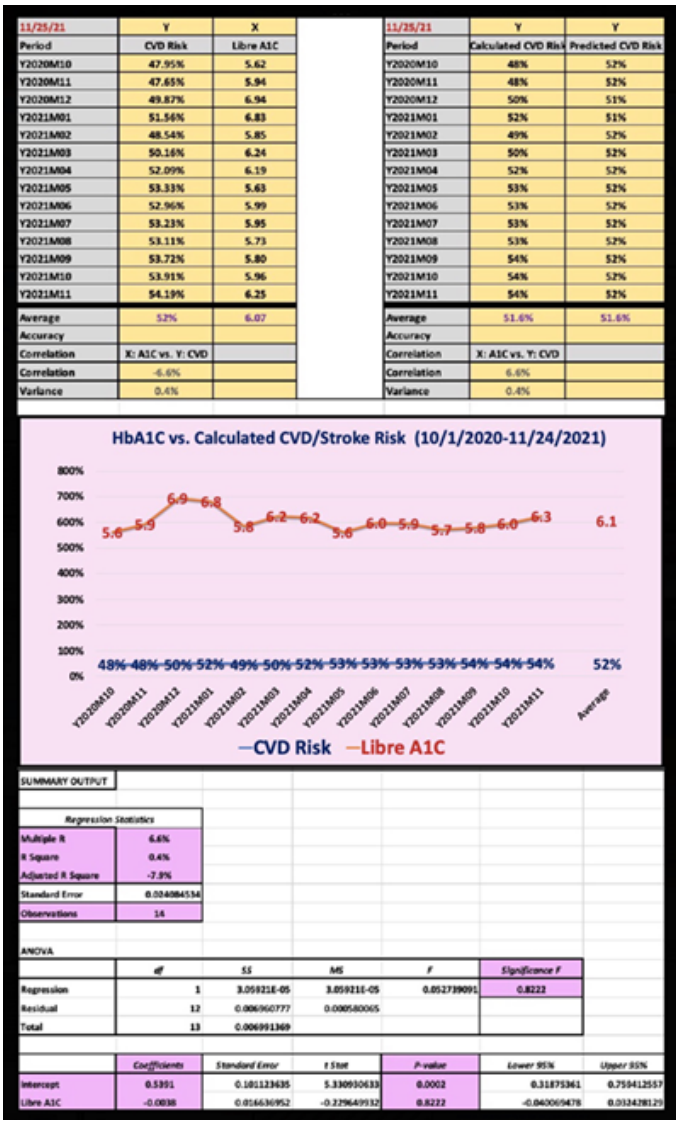
The author self-studied chronic diseases, metabolism, and food nutrition for 4-years from 2010 to 2013. He started his medical research work by building a mathematical metabolism model in 2014. He named his research methodology as the "GH-method: math-physical medicine (MPM approach)". Over the past 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 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 unhealthy 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 linear regression analyses of his CVD/Stroke risk versus the monthly HbA1C over a 14-month period from 10/1/2012 to 11/24/2021. There are 14 observations (months) with the significance F and p-value of 0.8222; therefore, the results are not statistically significant.



11/25/21	Y	X
Period	CVD Risk	Libre A1C
Y2020M10	47.95%	5.62
Y2020M11	47.65%	5.94
Y2020M12	49.87%	6.94
Y2021M01	51.56%	6.83
Y2021M02	48.54%	5.85
Y2021M03	50.16%	6.24
Y2021M04	52.09%	6.19
Y2021M05	53.33%	5.63
Y2021M06	52.96%	5.99
Y2021M07	53.23%	5.95
Y2021M08	53.11%	5.73
Y2021M09	53.72%	5.80
Y2021M10	53.91%	5.96
Y2021M11	54.19%	6.25
Average	52%	6.07
Accuracy		
Correlation	X: A1C vs. Y: CVD	
Correlation	-6.6%	
Variance	0.4%	

11/25/21	Y	Y
Period	Calculated CVD Risk	Predicted CVD Risk
Y2020M10	48%	52%
Y2020M11	48%	52%
Y2020M12	50%	51%
Y2021M01	52%	51%
Y2021M02	49%	52%
Y2021M03	50%	52%
Y2021M04	52%	52%
Y2021M05	53%	52%
Y2021M06	53%	52%
Y2021M07	53%	52%
Y2021M08	53%	52%
Y2021M09	54%	52%
Y2021M10	54%	52%
Y2021M11	54%	52%
Average	51.6%	51.6%
Accuracy		
Correlation	X: A1C vs. Y: CVD	
Correlation	6.6%	
Variance	0.4%	

SUMMARY OUTPUT

Regression Statistics	
Multiple R	6.6%
R Square	0.4%
Adjusted R Square	-7.9%
Standard Error	0.024084534
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	1	3.05921E-05	3.05921E-05	0.052739091	0.8222
Residual	12	0.006960777	0.000580065		
Total	13	0.006991369			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.5391	0.101123435	5.330930633	0.0002	0.31875361	0.759412557
Libre A1C	-0.0038	0.016636952	-0.229649932	0.8222	-0.040069478	0.032428129

Figure 1: Input data table and curves with regression analysis results of predicted CVD risk using HbA1C as input (10/1/2020 - 11/24/2021)

The 2 key data are listed as follows:

Correlation (R) = 6.6%
Variance (R^2) = 0.4%

Figure 2 illustrates the comparison of his regression predicted CVD risk (orange curve) versus the previously calculated CVD risk (blue curve) based on the MI model. It is noticeably clear that his regression predicted CVD risks are completely out-of-sync with the metabolism calculated CVD risks.

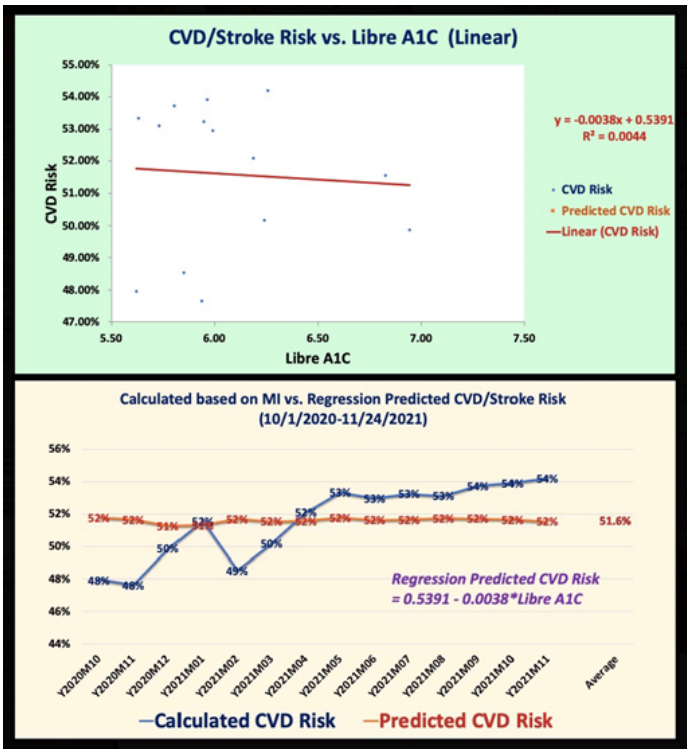


Figure 2: Results using HbA1C as input, with both space-domain scattered data diagram and time-domain curves of MI-calculated versus regression predicted CVD risks (10/1/2020 - 11/24/2021)

Figure 3 presents a summarized data table of the linear regression analyses of his CVD/Stroke risk versus the monthly combined medical conditions (weight m1, glucose m2, blood pressure m3, and lipids m4) over the same 14-month period from 10/1/2012 to 11/24/2021. There are 14 observations (months) with the significance F of 0.0000005 and p-value of 0.0000002; therefore, ***the results are statistically significant.***

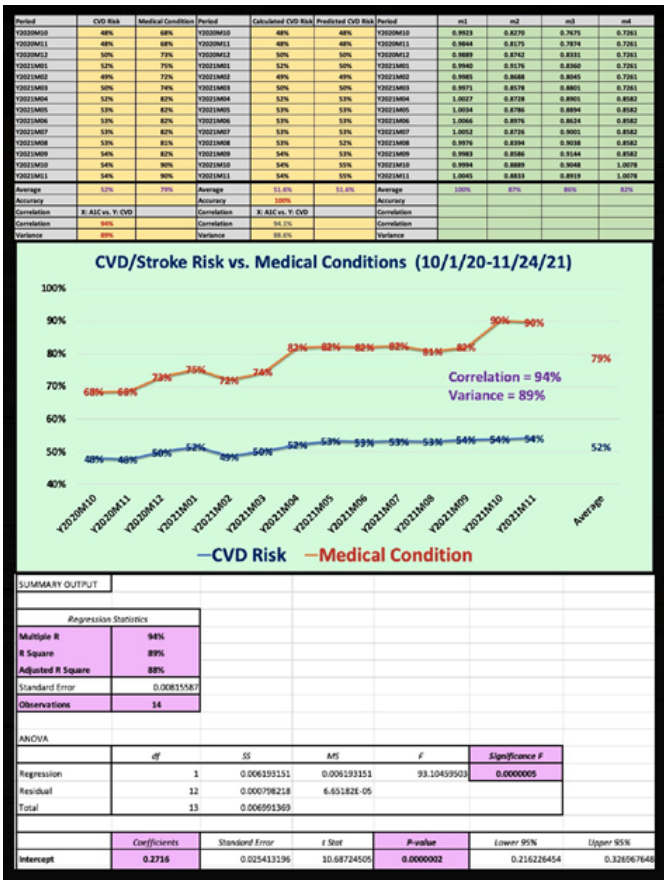


Figure 3: Input data table and curves with regression analysis results of predicted CVD risk using medical conditions as input (10/1/2020 - 11/24/2021)

The 2 key data are listed as follows:

Correlation (R) = 94%
Variance (R^2) = 89%

Figure 4 reflects the comparison of his regression predicted CVD risk (orange curve) versus the previously calculated CVD risk (blue curve) based on the MI model. It is evident that his regression predicted CVD risks are almost completely synchronizing with the MI calculated CVD risks.

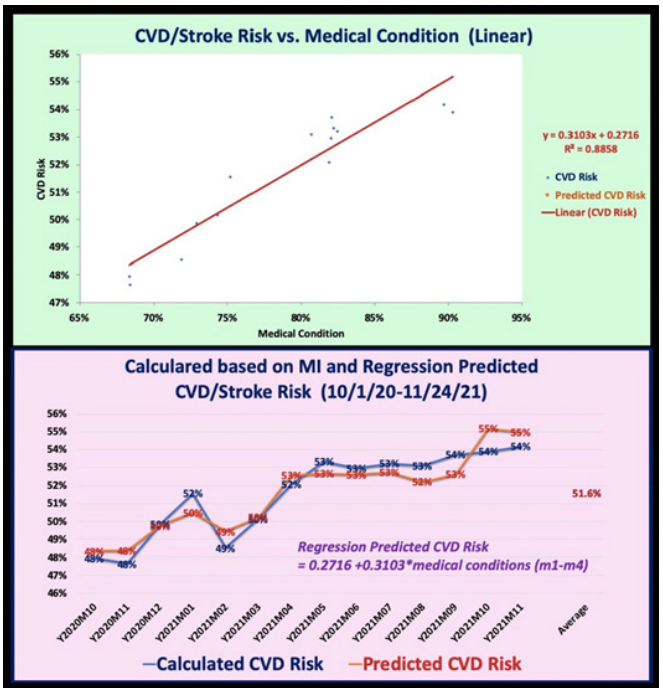
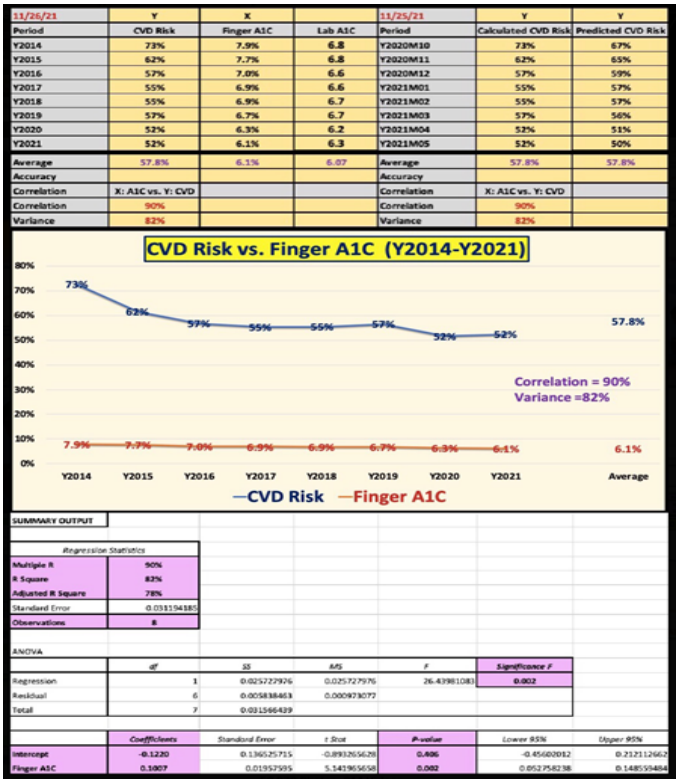


Figure 4: Results using medical conditions as input, with both space-domain scatter data diagram and time-domain curves of MI-calculated versus regression predicted CVD risks (10/1/2020 - 11/24/2021)

Figure 5 shows a summarized data table of the linear regression analyses of his CVD/Stroke risk versus the finger A1C values over the 8-year period from 2014 to 2021. There are 8 observations (years) with the significance F of 0.002 and p-value of 0.002; therefore, ***the results are statistically significant.***



11/26/21	Y	X		11/25/21	Y	Y
Period	CVD Risk	Finger A1C	Lab A1C	Period	Calculated CVD Risk	Predicted CVD Risk
Y2014	73%	7.9%	6.8	Y2020M10	73%	67%
Y2015	62%	7.7%	6.8	Y2020M11	62%	65%
Y2016	57%	7.0%	6.6	Y2020M12	57%	59%
Y2017	55%	6.9%	6.6	Y2021M01	55%	57%
Y2018	55%	6.9%	6.7	Y2021M02	55%	57%
Y2019	57%	6.7%	6.7	Y2021M03	57%	56%
Y2020	52%	6.3%	6.2	Y2021M04	52%	51%
Y2021	52%	6.1%	6.3	Y2021M05	52%	50%
Average	57.8%	6.1%	6.67	Average	57.8%	57.8%
Accuracy				Accuracy		
Correlation	X: A1C vs. Y: CVD			Correlation	X: A1C vs. Y: CVD	
Correlation	90%			Correlation	90%	
Variance	82%			Variance	82%	

SUMMARY OUTPUT

Regression Statistics	
Multiple R	90%
R Square	82%
Adjusted R Square	78%
Standard Error	0.031194185
Observations	8

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.025272926	0.025272926	26.43981083	0.002
Residual	6	0.005838463	0.000973077		
Total	7	0.031566439			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-0.1220	0.136525715	-0.893266628	0.406	-0.456002012	0.212112662
Finger A1C	0.1007	0.019577595	5.141965658	0.002	0.052758238	0.148559484

Figure 5: Input data table and curves with regression analysis results of predicted CVD risk using HbA1C as input (Y2014 - Y2021)

The 2 key data are listed as follows:

Correlation (R) = 90%
Variance (R^2) = 82%

Figure 6 demonstrates the comparison of his regression predicted CVD risk (orange curve) versus the previously calculated CVD risk (blue curve) based on the MI model. It is obvious that his regression predicted CVD risks are almost completely synchronizing with the MI calculated CVD risks.

Conclusions

In conclusion, his risk of having a CVD/Stroke is highly connected to the combined medical conditions, weight, glucose, blood lipids, blood pressure and heart rate, during the recent 14-month period. However, his risk of having a CVD/Stroke is not strongly related to the main diabetes measuring biomarker of HbA1C during the same period.

Of course, any analysis work using various statistical tools must pay attention on the section of the dataset and time period. Within two different time windows, the data distribution pattern and the data variability may differ according to the selected time-window; therefore, the analysis results and conclusions can vary.

The above observed two different conclusions based on the same 14-month period is due to the fact that his glucoses have been under stringent control (with an average HbA1C of 6.1%). Therefore, the variance during the 14-months using A1C as the input is a miniscule 0.4% compared to the variance using medical condition as input is an exceptionally high 89%.

If we examine the space-domain diagrams closely, using HbA1C as the input, the data results in the scattered map are spread out all over while a straight trend-line has a high difficulty to represent or simulate the majority of his CVD risk data. That is why its variance (R^2) is a mere 0.4%; therefore, the predicted CVD curve using A1C as the input is almost a horizontal line (similar to his A1C curve) and completely out-of-synch with the MI-model calculated CVD risk curve.

On the contrary, by using the medical condition as the input, the scattered results are located within a narrow data bend from the lower left corner to the upper right corner, whereas the straight trend-line represents 89% of the total CVD risk data. As a result, the predicted CVD curve using the medical condition as input is almost identical with the MI-model calculated CVD risk curve in the time-domain chart.

Furthermore, if he uses his finger A1C from the past 8 years (2014 to 2021), the scattered results are also located within a narrow data bend from the lower left corner to the upper right corner, while the straight trend-line represents 82% of the total CVD risk data. This conclusion is a result of the A1C curve trend which fluctuates and matches his CVD risk waveform from the past 8 years.

The findings from this article have demonstrated the importance of selection of both data and time-window which directly in-

fluence the final results and conclusions, if we uses statistical methods as our biomedical research tools.

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.