Applying the Distributional Data Analysis Tool, Biomarker Density, with the Collected Daily Data from 2 Types of Glucoses over the Past 3.5 Years of a Patient with Chronic Diseases to Investigate his Glucose Conditions Based on GH-Method: Math-Physical Medicine (No. 518)

Gerald C Hsu

EclaireMD Foundation, USA

*Corresponding author
Gerald C Hsu, EclaireMD Foundation, USA

Submitted: 09 Sep 2021; **Accepted**: 15 Sep 2021; **Published**: 28 Sep 2021

Abstract

Recently, the author conducted a series of medical research projects by applying a distributional data density analysis tool on his weight, glucose, blood pressure (BP), and heart conditions, while using his collected big data regarding certain biomarker's density distribution for the selected years.

In this article, he investigates his collected glucose data density via two different collection methods, finger-piercing and continuous glucose monitoring (CGM) sensor device, within a time span of 3.5 years (5/8/2018 - 9/13/2021).

With this data, he can interpret the results and explore additional and in-depth information since he is most familiar with his own health conditions. The findings from his own data is applicable to other patients of type 2 diabetes (T2D). The main purpose of writing this series of research articles is to further demonstrate the applicability and power of the specific distributional data density analysis tool.

*When he previously researched certain biomarkers and their relationships with other influential factors, he generally used the average values of those biomarkers. We know that most biomarkers, including glucoses, could fluctuate along the time scale in the form of a "wave". Each wave has its own unique amplitude and specific measuring unit which are associated with this particular biomarker. However, there are two other key factors, frequency and wavelength, to be considered as well. Particularly, the frequency component is associated with energy and excessive energy which causes damages to the internal organs. Therefore, without focusing on waveform of a biomarker and depending only on its mean value, we would lose many vital, interesting, and useful hidden information. This type of mean value, such as HbA1C, can only provide partial views of our overall diabetic conditions. These biomarkers still have missing information which carry certain hidden internal turmoil or vital signs, e.g. biomarker variation or its severe stimulation due to all types of external and/or internal stimulators. Therefore, by applying this basic knowledge of distributional data analysis by defining another term known as the "general biomarker density or **Bio-density%"** **(BMD%),** he can explore additional, different, in-depth, and useful hidden information from collected biomarker data and their associated waveforms.*

The term "biomarker density percentage" (BMD%) is defined as the occurrence frequency at a specific person's biomarker value. With this, he can calculate and examine each biomarker's occurrence rate within certain range over a selected timespan. This selected timespan is dependent on the study which is applied to specific patients (in this case, himself). *As of 1/1/2012 he started to track his daily weight and daily finger glucose and began col-* lecting his CGM sensor glucoses on 5/8/2018. As a result, his selected timespan for this particular study commenced on 5/8/2018 and ended on 9/13/2021. By investigating the changes of the peak biomarker value with their associated BD% from year to year, he can easily observe his biomarker's moving trend and understand his actual health problems or necessary health improvement effort clearly.

The above description provides the reason he keeps searching for applicable tools to analyze the collected big data of any biomarker. If this type of biomarker examination method is accepted by the medical community, it can be an extremely beneficial tool for doctors to quickly study the health conditions of their patients. Furthermore, the author programmed this algorithm into an iPhone APP software. Through the combination of his publishes papers and medical books along with a widely distributed APP for patient's use in the future, he believes that worldwide patients with chronic diseases can benefit from his research work. Hopefully, his research papers would not be limited within the scope of a "descriptive style using 26 alphabets" but instead as a "quantitative style using 10 digits". Numbers do not lie as long as we don't use fake, unorganized, and/or uncleaned data. Statistics is a tricky tool to use for any research work because it has the obvious characteristics of garbage in and garbage out (GIGO). It is also important to know that by using statistics with different selected time-windows for certain studies will result into varying conclusions.

In summary, *the author conducts this research work using the tools of Glucose Density % (GD%) with his collected 3 daily glucose: estimated average glucose (eAG), postprandial plasma glucose (PPG), and fasting plasma glucose (FPG) from the finger-piercing method and sensor collecting method over the same period of 3.5 years (5/8/2018 - 9/13/2021).*

Each of these 6 selected glucoses has its own unique glucose range (maximum glucose and minimum glucose) and specifically defined glucose "normal conditions", e.g. under 120 mg/dL is non-diabetes, whereas above 180 mg/dL is severe diabetes. This makes a combined study and data presentation quite difficult. In order to combine the 6 glucoses into one single diagram for this case, he must redefine a common "general-scale" of the glucose data range from 40 mg/dL to 260 mg/dL with an equal interval of 1 mg/dL. *With this new numbering system, he can then align these 6 different "normal conditions or target values" of glucoses based on the following definitions:*

Below the glucose level 70 mg/dL as "hypoglycemic";
Between 70 mg/dL and 180 mg/dL as "pre-diabetic or diabetic";
Above 180 mg/dL as "hyperglycemic".

For the purpose of additional investigation on the various "diabetic ranges", he has chosen a narrower dividing line of 140 mg/dL in order to distinguish the "pre-diabetes condition" and "diabetes condition".

Now, he is able to plot all 6 glucose density % curves into one combined diagram with their relative positions which indicate certain biomedical meanings. Through a closer examination of this combined diagram, he provides the following three conclusive statements:

(1) From the time-domain diagram and density-domain diagram, *his eAG curves and PPG curves are closer to each other*. This is logical since his finger PPG occupies 75% of weight for eAG (3 post-meal tests out of a total of 4 tests per day) and sensor PPG occupies 38% of weight for eAG (9 hours of PPG period versus 24 hours of eAG per day). On the other hand, *his FPG curves are distant from the curves of eAG and PPG*. This phenomenon

is a result from his finger FPG occupying only 25% of weight for eAG (1 fasting test out of a total of 4 tests per day) and sensor FPG occupies 29% of weight for eAG (7 hours of PPG period versus 24 hours of eAG per day). The other missing percentages belong to the glucoses associated with between-meals and pre-bed time periods, 0% for finger case (no test conducted) and 33% for sensor case (8 hours of 24 hours).

(2) Examining his time-in-range *(TIR) percentages, it is clear that his finger glucose densities within the TIR range are higher than the sensor glucose densities within the same TIR range*. In the case of finger versus sensor between 70 mg/dL - 120 mg/dL, *78%>53% for eAG, 77%>44% for PPG, 83%>79% for FPG. In the case of finger versus sensor between 70 mg/dL - 140 mg/dL, 97%>78% for eAG, 96%>73% for PPG, 99%>92% for FPG. In the case of finger versus sensor between 70 mg/dL - 180 mg/dL, 99%>96% for eAG, 99%>96% for PPG, 99%>92% for FPG.*

(3) With hypoglycemia (<70 mg/dL) and hyperglycemia (>180 mg/dL), his case demonstrates extremely low probabilities for either one. Actually, his TIR percentages for 70 mg/dL - 180 mg/dL indicate that his scores are 99% for finger glucoses and 97% for sensor glucoses. This means that his T2D control has been remarkably effective during the period from 5/8/2018 to 9/13/2021.

By combining the two different analysis methods, the traditional time-domain analysis and the newly defined density-domain analysis, he can explore more insights of his glucoses.
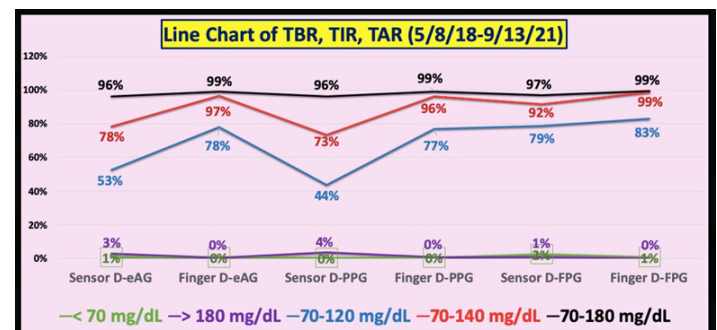
Introduction
Recently, the author conducted a series of medical research projects by applying a distributional data density analysis tool on his weight, glucose, blood pressure (BP), and heart conditions, while using his collected big data regarding certain biomarker's density distribution for the selected years.

In this article, he investigates his collected glucose data density via two different collection methods, finger-piercing and continuous glucose monitoring (CGM) sensor device, within a time span of *3.5 years (5/8/2018 - 9/13/2021).*

With this data, he can interpret the results and explore additional and in-depth information since he is most familiar with his own health conditions. The findings from his own data is applicable to other patients of type 2 diabetes (T2D). The main purpose of writing this series of research articles is to further demonstrate the applicability and power of the specific distributional data *density* analysis tool.

When he previously researched certain biomarkers and their relationships with other influential factors, he generally used the average values of those biomarkers. We know that most biomarkers, including glucoses, could fluctuate along the time scale in the form of a "wave". Each wave has its own unique amplitude and specific measuring unit which are associated with this particular biomarker. However, there are two other key factors, frequency and wavelength, to be considered as well. Particularly, the frequency component is associated with energy and excessive energy which causes damages to the internal organs. Therefore, without focusing on waveform of a biomarker and depending only on its mean value, we would lose many vital, interesting, and useful hidden information. This type of mean value, such as HbA1C, can only provide partial views of our overall diabetic conditions. These biomarkers still have missing information which carry certain hidden internal turmoil or vital signs, e.g. biomarker variation or its severe stimulation due to all types of external and/or internal stimulators. Therefore, by applying this basic knowledge of distributional data analysis by defining another term known as the "***general biomarker density or Bio-density%" (BMD%)***, he can explore additional, different, in-depth, and useful hidden information from collected biomarker data and their associated waveforms.

The term "biomarker density percentage" (BMD%) is defined as the occurrence frequency at a specific person's biomarker value. With this, he can calculate and examine each biomarker's occurrence rate within certain range over a selected timespan. This selected timespan is dependent on the study which is applied to specific patients (in this case, himself). *As of 1/1/2012 he started to track his daily weight and daily finger glucose and began collecting his CGM sensor glucoses on 5/8/2018. As a result, his selected timespan for this particular study commenced on 5/8/2018 and ended on 9/13/2021.* By investigating the changes of the peak biomarker value with their associated BD% from year to year, he can easily observe his biomarker's moving trend and understand his actual health problems or necessary health improvement effort clearly.

The above description provides the reason he keeps searching for applicable tools to analyze the collected big data of any biomarker. If this type of biomarker examination method is accepted by the medical community, it can be an extremely beneficial tool for doctors to quickly study the health conditions of their patients. Furthermore, the author programmed this algorithm into an iPhone APP software. Through the combination of his publishes papers and medical books along with a widely distributed APP for patient's use in the future, he believes that worldwide patients with chronic diseases can benefit from his research work. Hopefully, his research papers would not be limited within the scope of a "descriptive style using 26 alphabets" but instead as a "quantitative style using 10 digits". Numbers do not lie as long as we don't use fake, unorganized, and/or uncleaned data. Statistics is a tricky tool to use for any research work because it has the obvious characteristics of garbage in and garbage out (GIGO). It is also important to know that by using statistics with different selected time-windows for certain studies will result into varying conclusions.

Methods
MPM Background
To learn more about his developed GH-Method: math-physical medicine (MPM) methodology, readers can read the following three papers selected 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, his 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 T2D patient since 1996. He weighed 220 lb. (100 kg, BMI 32.5) at that time. By 2010, he still weighed 198 lb. (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 from 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, 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 heavier traveling period.

During 2020 with a COVID-19 quarantined lifestyle, not only has he 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 his knowledge of chronic diseases, practical lifestyle management experiences, and developed various high-tech tools contribute to his excellent health status since 1/19/2020, which is the start date of being self-quarantined.

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). By the way, 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 2+ 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, using optical physics, artificial intelligence (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 chronic kidney disease (CKD), bladder, foot, and eye issues such as diabetic retinopathy (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, semantic, and COVID-19.

To date, he has collected more than two million data regarding his medical conditions and lifestyle details. In addition, he has written 498 medical papers and published 400+ articles in 100+ various medical journals, including 6 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 and shared his findings and learnings with other patients worldwide.

Glucose Density (GD)
For the case of one particular patient i, the collected biomarker data can be expressed by pairs of data in the format of (t_{ij}, X_{ij}), $j = 1 \ldots T$, where the t_{ij} represent recording times and X_{ij} is the biomarker level at time instant t_{ij}, and T is the overall observation length of the selected biomarker. *For the case in this article, the total T is 221 (e.g. from 40 mg/dL to 260 mg/dL with an equal interval of 1 mg/dL between two glucose end-points).*

Therefore, he can describe the above mathematical problem into a more simplified equation for one patient only. The glucose density *% (GD%)* for one patient can be defined in terms of a *continuous format* as follows:

$$GD(x) = \int_{1}^{T} (Y(t)\, dt) / T$$

with $x1 < Y(t) < x2$
where x1 and x2 are $\int$ boundaries of his selected glucose range.

The glucose density % (GD%) equation for one patient, such as himself, can also be defined in terms of a discrete format as follows:

$$GD(x) = \left(\sum_{j=1}^{T} Y(tj) \right) / T$$

with $x1 < Y(t) < x2$
where x1 and x2 are boundaries of his selected glucose range.

He then develops his APP software program using the above-described algorithm.

Results
Figure 1 shows the comparison between his finger glucoses and his CGM sensor glucoses using a tome-domain analysis. It should be pointed out that the *correlations between the finger curves and sensor cures are within a high range of 77% - 89%.*
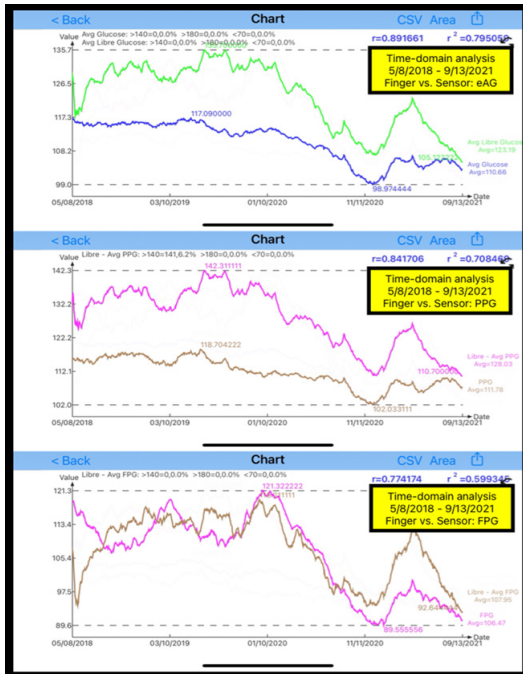
Figure 1: Time-domain analysis of his eAG, PPG, FPG curves within a period of 3.5 years (5/8/2018 - 9/13/2021)

Figure 2 illustrates the comparison between finger glucoses and sensor glucoses along with eAG, PPG, and FPG using a density-domain analysis. The 3 diagrams are generated on his APP with different glucose ranges that contain varying minimum glucoses and maximum glucoses; however, they have *similar patterns between eAG and PPG as well as different patterns between FPG and eAG/PPG which are evident.*
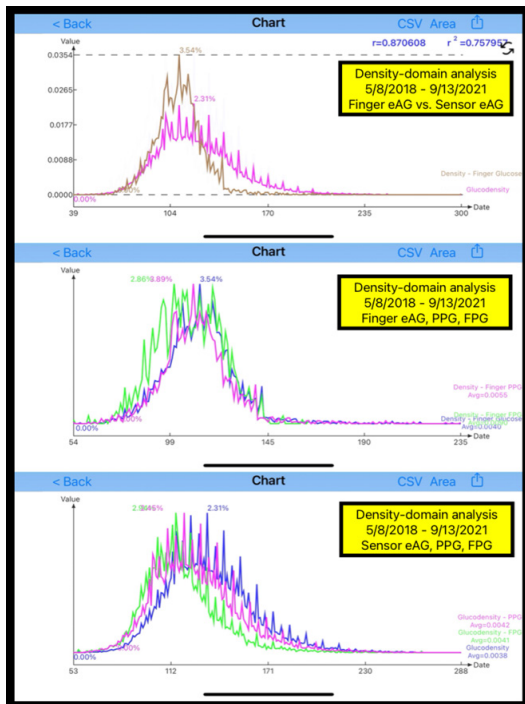

Figure 2: Density-domain analysis using 6 glucose data via APP

software within a period of 3.5 years (5/8/2018 - 9/13/2021)

Figure 3 reflects the combined glucose densities of the 6 glucoses into one diagram which has consistent glucose range from 40 mg/dL to 260 mg/dL. *Again, they have similar patterns between eAG and PPG as well as different patterns between FPG and eAG/PPG which are obvious.* Furthermore, the higher GD peaks are associated with the Finger glucoses compared to the GD peaks of Sensor glucoses that can be found if we examine them closely from this combined figure.
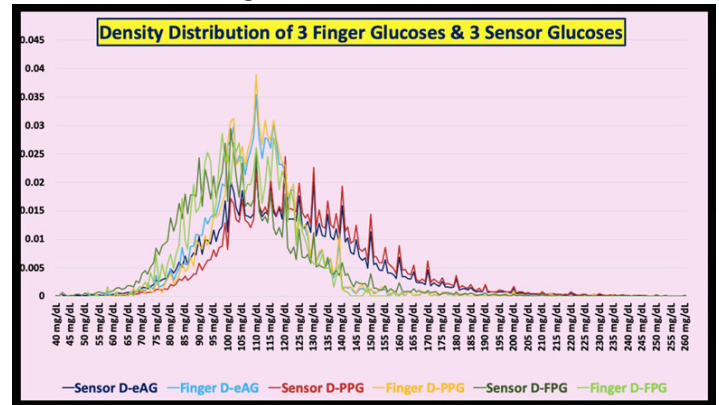

Figure 3: A combined glucose density GD% distribution diagram of 6 selected glucoses based on the same data range within a common period of 3.5 years (5/8/2018 - 9/13/2021)

Figure 4 uses the line chart and bar chart to display more hidden information from the above glucose density diagrams. By examining the *TIR percentages, it is clear that his finger glucose densities within the TIR range are higher than the sensor glucose densities within the same TIR range.*
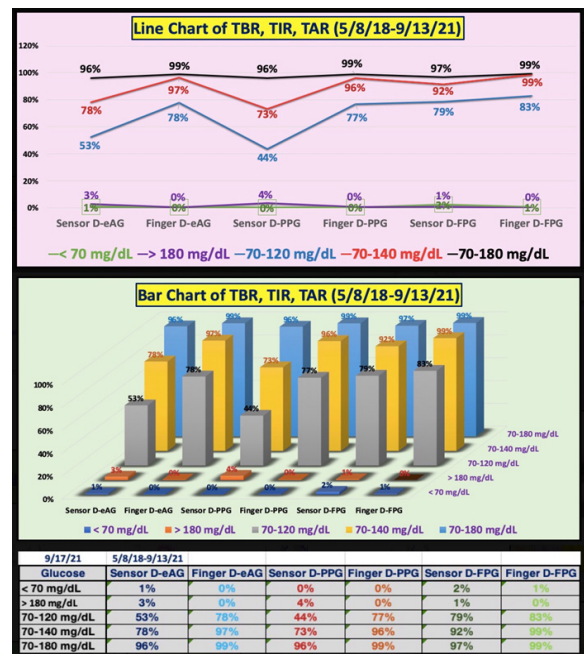

Figure 4: TBR/TIR/TAR analysis using combined glucose density GD% distribution of 6 selected glucoses within a period of 3.5

years (5/8/2018 - 9/13/2021)

In the case of finger versus sensor between 70 mg/dL - 120 mg/dL, *78%>53% for eAG, 77%>44% for PPG, 83%>79% for FPG*.

In the case of finger versus sensor between 70 mg/dL - 140 mg/dL, *97%>78% for eAG, 96%>73% for PPG, 99%>92% for FPG*.

In the case of finger versus sensor between 70 mg/dL - 180 mg/dL, *99%>96% for eAG, 99%>96% for PPG, 99%>92% for FPG*.

Furthermore, with hypoglycemia (<70 mg/dL) and hyperglycemia (>180 mg/dL), his case demonstrates having extremely low probabilities. Actually, both his TIR percentages for 70 mg/dL - 180 mg/dL have indicated that his scores are 99% for finger glucoses and 97% for sensor glucoses. This means that *his T2D control has been very effective during the period from 5/8/2018 to 9/13/2021.*

Conclusions

In summary, the author conducts this research work using the tools of *Glucose Density % (GD%)* with his collected *3 daily glucose: estimated average glucose (eAG), postprandial plasma glucose (PPG), and fasting plasma glucose (FPG) from the finger-piercing method and sensor collecting method over the same period of 3.5 years (5/8/2018 - 9/13/2021).*

Each of these 6 selected glucoses has its own unique glucose range (maximum glucose and minimum glucose) and specifically defined glucose "normal conditions", e.g. under 120 mg/dL is non-diabetes, whereas above 180 mg/dL is severe diabetes. This makes a combined study and data presentation quite difficult. In order to combine the 6 glucoses into one single diagram for this case, he must redefine a common "general-scale" of the glucose data range from 40 mg/dL to 260 mg/dL with an equal interval of 1 mg/dL. *With this new numbering system, he can then align these 6 different "normal conditions or target values" of glucoses based on the following definitions:*

Below the glucose level 70 mg/dL as "hypoglycemic";
Between 70 mg/dL and 180 mg/dL as "pre-diabetic or diabetic";
Above 180 mg/dL as "hyperglycemic".

For the purpose of additional investigation on the various "diabetic ranges", he has chosen a narrower dividing line of 140 mg/dL in order to distinguish the "pre-diabetes condition" and "diabetes condition".

Now, he is able to plot all 6 glucose density % curves into one combined diagram with their relative positions which indicate certain biomedical meanings. Through a closer examination of this combined diagram, he provides the following three conclusive statements:

(1) From the time-domain diagram and density-domain diagram*, his eAG curves and PPG curves are closer to each other.* This is logical since his finger PPG occupies 75% of weight for eAG (3 post-meal tests out of a total of 4 tests per day) and sensor PPG occupies 38% of weight for eAG (9 hours of PPG period versus 24 hours of eAG per day). On the other hand*, his FPG curves are distant from the curves of eAG and PPG.* This phenomenon is a result from his finger FPG occupying only 25% of weight for eAG (1 fasting test out of a total of 4 tests per day) and sensor FPG occupies 29% of weight for eAG (7 hours of PPG period versus 24 hours of eAG per day). The other missing percentages belong to the glucoses associated with between-meals and pre-bed time periods, 0% for finger case (no test conducted) and 33% for sensor case (8 hours out of 24 hours).

(2) Examining his time-in-range *(TIR) percentages, it is clear that his finger glucose densities within the TIR range are higher than the sensor glucose densities within the same TIR range.* In the case of finger versus sensor between 70 mg/dL - 120 mg/dL, *78%>53% for eAG, 77%>44% for PPG, 83%>79% for FPG.* In the case of finger versus sensor between 70 mg/dL - 140 mg/dL, *97%>78% for eAG, 96%>73% for PPG, 99%>92% for FPG*. In the case of finger versus sensor between 70 mg/dL - 180 mg/dL, *99%>96% for eAG, 99%>96% for PPG, 99%>92% for FPG.*

(4) With hypoglycemia (<70 mg/dL) and hyperglycemia (>180 mg/dL), his case demonstrates extremely low probabilities for either one. Actually, his TIR percentages for 70 mg/dL - 180 mg/dL indicate that his scores are 99% for finger glucoses and 97% for sensor glucoses. This means that his T2D control has been remarkably effective during the period from 5/8/2018 to 9/13/2021.

By combining the two different analysis methods, the traditional time-domain analysis and the newly defined density-domain analysis, he can explore more insights of his glucoses.

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.

1. Marcos Matabuena1, Alexander Petersen, Juan C Vidal, Francisco Gude (2020) Glucodensities: a new representation of glucose profiles using distributional data analysis https://arxiv.org/pdf/2008.07840.pdf