ISSN: 2689-1204

Research Article

Journal of Applied Material Science & Engineering Research

Illustration of Finger-Piercing Glucose Data Density Study for a Period of 8.6 Years from 1/1/2012 to 9/18/2021 of a type 2 Diabetes Patient Using Both Statistical Gaussian or Normal Distribution model and Biomedical TxR Analysis Model as Comparison Tools Based on GH-Method: Math-Physical Medicine (No. 520)

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

Citation: *Gerald C Hsu (2021) Illustration of Finger-Piercing Glucose Data Density Study for a Period of 8.6 Years from 1/1/2012 to 9/18/2021 of a type 2 Diabetes Patient Using Both Statistical Gaussian or Normal Distribution model and Biomedical TxR Analysis Model as Comparison Tools Based on GH-Method: Math-Physical Medicine (No. 520). J App Mat Sci & Engg Res, 5(3), 1-5.*

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 different timespans.

With his personal collected data, he can interpret the results and explore additional information since he is most familiar with his own health conditions. Of course, these findings regarding his own body is also applicable to other patients with chronic diseases. The main purpose of writing this series of research articles is to further demonstrate the applicability and power of using this specific distributional data density analysis tool.

In the past, when he researched certain biomarkers and their relationships with other influential factors, he generally used the average values of those biomarkers. However, we know that most biomarkers, including body weight, glucoses and BP, would fluctuate along the time scale in the form of a "wave". Each wave has its unique amplitude and specific biomedical measuring unit which are associated with this particular biomarker. There are two other key factors, frequency and wavelength, which need to be considered as well. Particularly, the frequency component is associated with energy and excessive energy carried by blood cells to circulate inside of body would cause damages to our 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. These types of mean values, such as HbA1C, or sparsely collected finger-pierced glucose, or quarterly available lab-tested blood lipid results, can only provide partial views of the overall health conditions. Those biomarkers still have some missing information that carry hidden internal turmoil or vital signs, e.g. biomarker variation or its severe stimulation due to all types of external and internal stimulators. By applying this basic knowledge of distributional data density analysis by defining a new term known as the "general biomarker density or Bio-density%" (BMD%), he can explore additional, different, deeper, and useful hidden information from the collected biomarker data and their associated waveforms.

The term "biomarker density percentage" (BMD%) is defined as the occurrence frequency at a specific biomarker value. With this, he can calculate and examine each biomarker's occurrence rate within a certain range over his selected timespan. This selected timespan is dependent on the study which is applicable to specific patients (in this article, himself). By investigating the changes of the peak biomarker value with their associated BMD% from year to year, he can easily observe the 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 has programmed this algorithm into an iPhone-based APP software. Through the combination of his pub-

lishes 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 data, 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.

The author has tracked his daily finger-piercing glucose since 1/1/2012; therefore, his selected timespan for this study is 8.6 years starting from 1/1/2012 and ending on 9/18/2021. In this particular article, he investigates his collected finger-piercing glucose data density curve within a time span of 8.6 years (1/1/2012 - 9/18/2021) to compare it against the statistical Gaussian Distribution (or Normal distribution) model to check their degree of curve similarity.

In summary, the author wrote this article using the GD% results based on his collected daily glucoses via finger-piercing method over a period of 8.6 year, from 1/1/2012 to 9/18/2021. He then compares his calculated GD% results with the statistical Gaussian Distribution (or Normal Distribution) model and the biomedical TxR (i.e. TBR/TIR/TAR) model defined by American Diabetes Association.

The selected glucoses have their own unique glucose range, maximum glucose and minimum glucose, and certain specifically defined glucose "normal biomedical conditions" in medical community, for example, glucose values under 120 mg/dL is non-diabetes, above 180 mg/dL is severe diabetes, below 70 mg/dL is hypoglycemia, and above 180 mg/dL is hyperglycemia. His generated GD% curve through his APP covers a wide range from 51 mg/dL to 300 mg/dL. However, in reality, for those glucoses above 180 mg/dL have very small quantity of data. Therefore, he has decided to retain his glucose data within the range of 50 mg/dL and 180 mg/dL, and truncated off those small-quantity data from his GD% diagram. With this modification, his GD% curves have shown visually an extremely high similarity to a standard statistical Gaussian Distribution curve.

Furthermore, he utilized the following 3 diabetes definitions as another set of biomedical comparison:

(1) Below the glucose level 70 mg/dL as "hypoglycemic" or "time-below-range TBR".
(2) The glucose level between 70 mg/dL and 180 mg/dL as "normal range" or "time-in-range TIR".
(3) Above the glucose level 180 mg/dL as "hyperglycemic" or "time-above-range TAR".

His results have demonstrated the following statistical conclusions using the standard deviation (SD) in a format of (calculate GD%, imperial rule's % for a standard normal distribution) for comparison.

Within 1SD range: (47%, 68%)
Within 2SD range: (72%, 95%)
Within 3SD range: (89%, 99.7%)

Additionally, his diabetes analysis results using ADA's TxR model have obtained another set of comparison conclusions using the diabetes biomedical standards of TAR, TIR, and TAR:

TBR: 0.3%
TIR: 97.1%
TAR: 2.5%

As a result, ***his risk of having hypoglycemia (insulin shock) are negligible and his risk of having hyperglycemia (severely high glucoses) are also extremely low.***

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 different timespans.

With his personal collected data, he can interpret the results and explore additional information since he is most familiar with his own health conditions. Of course, these findings regarding his own body is also applicable to other patients with chronic diseases. The main purpose of writing this series of research articles is to further demonstrate the applicability and power of using this specific distributional data density analysis tool.

In the past, when he researched certain biomarkers and their relationships with other influential factors, he generally used the average values of those biomarkers. However, we know that most biomarkers, including body weight, glucoses and BP, would fluctuate along the time scale in the form of a "wave". Each wave has its unique amplitude and specific biomedical measuring unit which are associated with this particular biomarker. There are two other key factors, frequency and wavelength, which need to be considered as well. Particularly, the frequency component is associated with energy and excessive energy carried by blood cells to circulate inside of body would cause damages to our 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. These types of mean values, such as HbA1C, or sparsely collected finger-pierced glucose, or quarterly available lab-tested blood lipid results, can only provide partial views of the overall health conditions. Those biomarkers still have some missing information that carry hidden internal turmoil or vital signs, e.g. biomarker variation or its severe stimulation due to all types of external and internal stimulators. By applying this basic knowledge of distributional data ***density*** analysis by defining a new term known as the "***general biomarker density or Bio-density%" (BMD%),*** he can explore additional, different, deeper, and useful hidden information from the collected biomarker data and their associated waveforms.

The term "biomarker density percentage" (BMD%) is defined as the occurrence frequency at a specific biomarker value. With this, he can calculate and examine each biomarker's occurrence rate within a certain range over his selected timespan. This select-

ed timespan is dependent on the study which is applicable to specific patients (in this article, himself). By investigating the changes of the peak biomarker value with their associated BMD% from year to year, he can easily observe the 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 has programmed this algorithm into an iPhone-based 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 data, 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.

The author has tracked his daily finger-piercing glucose since 1/1/2012; therefore, his selected timespan for this study is 8.6 years starting from 1/1/2012 and ending on 9/18/2021. In this particular article, he investigates his collected finger-piercing glucose data density curve within a time span of 8.6 years (1/1/2012 - 9/18/2021) to compare it against the statistical Gaussian Distribution (or Normal distribution) model to check their degree of curve similarity.

Methods

MPM Background

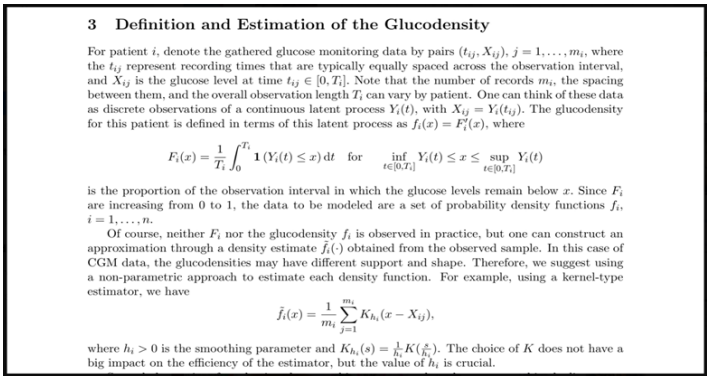
To learn more about his developed GH-Method: math-physical medicine (MPM) methodology, readers can read the following three papers selected from his ~500 published medical papers.

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".

Glucose Density (GD%)

The author took the following photo directly from the beginning part of Section 3 in the Glucodensities' paper (Reference 1), because he does not know how to write articles with LATAX math symbols on his iPAD Page application.



3 Definition and Estimation of the Glucodensity

For patient i, denote the gathered glucose monitoring data by pairs (t_{ij}, X_{ij}), $j = 1, \ldots, m_i$, where the t_{ij} represent recording times that are typically equally spaced across the observation interval, and X_{ij} is the glucose level at time $t_{ij} \in [0, T_i]$. Note that the number of records m_i, the spacing between them, and the overall observation length T_i can vary by patient. One can think of these data as discrete observations of a continuous latent process $Y_i(t)$, with $X_{ij} = Y_i(t_{ij})$. The glucodensity for this patient is defined in terms of this latent process as $f_i(x) = F_i'(x)$, where

$$F_i(x) = \frac{1}{T_i}\int_0^{T_i} \mathbf{1}\left(Y_i(t) \le x\right) \mathrm{d}t \quad \text{for} \quad \inf_{t\in[0,T_i]} Y_i(t) \le x \le \sup_{t\in[0,T_i]} Y_i(t)$$

is the proportion of the observation interval in which the glucose levels remain below x. Since F_i are increasing from 0 to 1, the data to be modeled are a set of probability density functions f_i, $i = 1, \ldots, n$.

Of course, neither F_i nor the glucodensity f_i is observed in practice, but one can construct an approximation through a density estimate $\hat{f}_i(\cdot)$ obtained from the observed sample. In this case of CGM data, the glucodensities may have different support and shape. Therefore, we suggest using a non-parametric approach to estimate each density function. For example, using a kernel-type estimator, we have

$$\hat{f}_i(x) = \frac{1}{m_i}\sum_{j=1}^{m_i} K_{h_i}(x - X_{ij}),$$

where $h_i > 0$ is the smoothing parameter and $K_{h_i}(s) = \frac{1}{h_i}K(\frac{s}{h_i})$. The choice of K does not have a big impact on the efficiency of the estimator, but the value of h_i is crucial.

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 … 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 131 (e.g. from 51 mg/dL to 181 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 ∫ 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.

Statistical Normal Distribution

There are over 20 different types of data distributions (applied to the continuous or the discrete space) commonly used in data science to model various types of phenomena. Normal distribution is one of the most important cases. There are infinitely many types of "normal distributions", but only one "Standard Normal Distribution".

In the case of normal distribution, the standard deviation (SD) and the mean together can tell us where most of the values in our statistical distribution lie if they follow a normal distribution. The ***empirical rule***, or the "***68-95-99.7" rule***, tells us where our values lie in the standard normal distribution case:

1. Around 68% of scores are within 1 standard deviation of the mean,
2. Around 95% of scores are within 2 standard deviations of the mean,
3. Around 99.7% of scores are within 3 standard deviations of the mean.

Results

Figure 1 shows a sample diagram of his 9 GD% curves during 2020. In the top area, the author has a collection of his 8 randomly selected monthly GD% curves, while the bottom part depicts the annual GD% curve of Y2020.

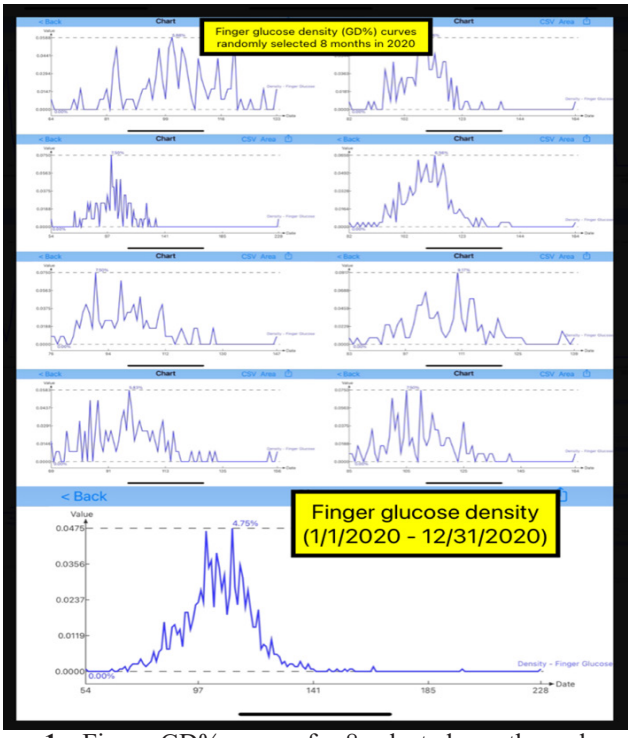


Figure 1: Finger GD% curves for 8 selected months and annual 2012

We can see that all of his 8 monthly GD% curves have different forms of data distribution patterns, but the majority of them still have a reasonable similarity with a normal distribution curve. However, the annual GD% curve in the bottom section looks most similar to a normal distribution curve.

Figure 2 is his GD% distribution curve for a longer period of 8.6 years from 1/1/12 to 9/18/2021. This 8.6-years distribution curve's shape is remarkably close to a normal distribution except for the width of its "peak lump" is wider than the "peak lump" of Y2020. This indicates that a wider range of glucoses have been collected over the 8.6-years long period compared to Y2020 of much shorter one-year period with a "controlled diabetes" situation.

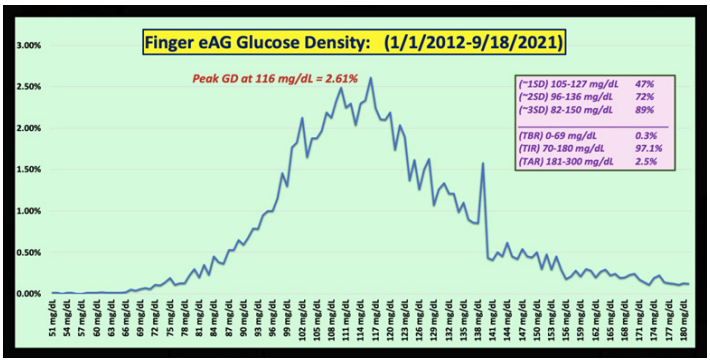


Figure 2: Finger GD% curve for a long period of 8.6 years (1/1/2012 - 9/18/2021) and Comparison with a normal distribution model

Furthermore, after running some more calculations, he has obtained the following two sets of data for further comparison.

First, by using the standard deviation (SD) calculation, he has obtained the following statistical results expressed in a format of (calculate GD%, imperial rule's % for a standard normal distribution):

Within 1SD range: (47%, 68%)
Within 2SD range: (72%, 95%)
Within 3SD range: (89%, 99.7%)

Although his GD% does not follow the empirical rule of a "standard normal distribution" exactly, it still contains the basic characteristics of a general "Gaussian distribution or normal distribution" pattern using 1SD, 2SD, and 3SD ranges.

Second, by utilizing the diabetes biomedical standards of TAR, TIR, and TAR, he discovered the following three more different biomedical results for diabetes control over the selected 8.6-years period:

TBR: 0.3%
TIR: 97.1%
TAR: 2.5%

His risk of having hypoglycemia (insulin shock) are negligible and his risk of having hyperglycemia (severely high glucoses) are also extremely low. This means that ***his T2D control has been quite effective during the overall period of 8.6-years from 1/1/2012 to 9/18/2021***.

Figure 3 illustrates his Finger GD% comparison between two extreme-ending years of Y2012 (1/1/2012 - 12/31/2012) and Y2021 (1/1/2021 - 9/18/2021). By the way, both of his two GD% curves are similar to the normal Gaussian distribution or normal distribution model.

The following table lists the comparison results of 3 different periods using a format of (TBR, TIR, and TAR) values:

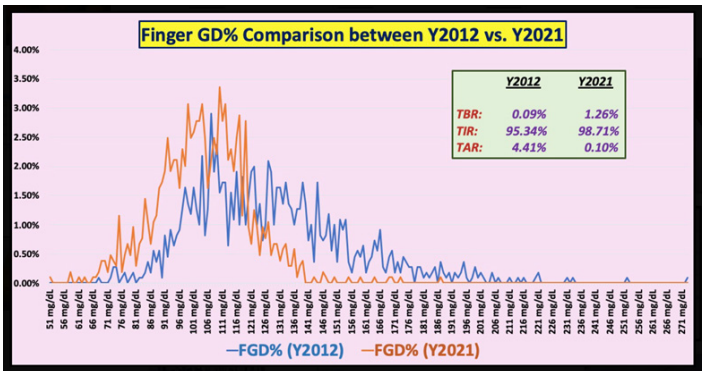


Figure 3: ADA's TxR (TBR, TIR, TAR) analysis results using GD% curves of 2012 vs. 2021

Y2012: (0.1%, 95.3%, 4.4%)
Y2012-Y2021: (0.3%, 97.1%, 2.5%)
Y2021: (1.3%, 98.7%, 0.1%)

The orders of TxR values for the above 3 selected periods have indicated that ***his diabetes control situations with Y2021 is the best, 8.6-years is the medium, and Y2012 is the worst. It should also be pointed out that his TBR in Y2021 is the highest due to his overall lower glucoses resulted from better control of his diabetes in Y2021, but his damaged pancreatic beta cells health situations are still existing.***

Conclusions

In summary, the author wrote this article using the GD% results based on his collected daily glucoses via finger-piercing method over a period of 8.6 year, from 1/1/2012 to 9/18/2021. He then compares his calculated GD% results with the statistical Gaussian Distribution (or Normal Distribution) model and the biomedical TxR (i.e. TBR/TIR/TAR) model defined by American Diabetes Association.

The selected glucoses have their own unique glucose range, maximum glucose and minimum glucose, and certain specifically defined glucose "normal biomedical conditions" in medical community, for example, glucose values under 120 mg/dL is non-diabetes, above 180 mg/dL is severe diabetes, below 70 mg/dL is hypoglycemia, and above 180 mg/dL is hyperglycemia. His generated GD% curve through his APP covers a wide range from 51 mg/dL to 300 mg/dL. However, in reality, for those glucoses above 180 mg/dL have very small quantity of data. Therefore, he has decided to retain his glucose data within the range of 50 mg/dL and 180 mg/dL, and truncated off those small-quantity data from his GD% diagram. With this modification, his GD% curves have shown visually an extremely high similarity to a standard statistical Gaussian Distribution curve.

Furthermore, he utilized the following 3 diabetes definitions as another set of biomedical comparison:

1. Below the glucose level 70 mg/dL as "hypoglycemic" or "time-below-range TBR".
2. The glucose level between 70 mg/dL and 180 mg/dL as "normal range" or "time-in-range TIR".
3. Above the glucose level 180 mg/dL as "hyperglycemic" or "time-above-range TAR".

His results have demonstrated the following statistical conclusions using the standard deviation (SD) in a format of (calculate GD%, imperial rule's % for a standard normal distribution) for comparison.

Within 1SD range: (47%, 68%)
Within 2SD range: (72%, 95%)
Within 3SD range: (89%, 99.7%)

Additionally, his diabetes analysis results using ADA's TxR model have obtained another set of comparison conclusions using the diabetes biomedical standards of TAR, TIR, and TAR:

TBR: 0.3%
TIR: 97.1%
TAR: 2.5%

As a result, *his risk of having hypoglycemia (insulin shock) are negligible and his risk of having hyperglycemia (severely high glucoses) are also extremely low.*

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