



Article

# Developing a Health Risk Evaluation Method for Triple H

Chien-Chih Wang <sup>1,\*</sup> , Cheng-Ding Chang <sup>2</sup> and Bernard C. Jiang <sup>3</sup>

<sup>1</sup> Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei 24301, Taiwan

<sup>2</sup> Department of Industrial Engineering and Management, Yuan Ze University, Chung-Li 32003, Taiwan; Justin\_Chang@unimicron.com

<sup>3</sup> Department of Industrial Management, National Taiwan University of Science and Technology, Taipei 10607, Taiwan; bcjiang@mail.ntust.edu.tw

\* Correspondence: iecchwang@mail.mcut.edu.tw; Tel.: +886-2-2908-9899

Received: 24 February 2019; Accepted: 30 March 2019; Published: 1 April 2019



**Abstract:** The development of a health evaluation system from human-related data is an important issue in preventive medicine. Previously, most studies have focused on disease assessment and prevention in patients. However, even if certain risk factors are all within normal ranges, individuals may not necessarily be completely healthy. This study focused on healthy individuals to develop a new index to assess health risks; this index can be used for the prevention of multiple diseases in healthy people. The kernel density technique was proposed to estimate the distribution of common risk factors and to develop a health risk index. A dataset of hypertension, hyperlipidemia, and hyperglycemia (Triple H) data from the National Health Insurance Research Database in Taiwan was used to demonstrate the proposed analytical process. The results of risk factor changes after six weeks of exercise were used to calculate the health risk index. The results showed that the subjects experienced a 7.29% reduction in their health risk index after the exercise intervention. This finding demonstrates the potential impact of an important reference index on quantifying the effect of maintenance in healthy people.

**Keywords:** multi-diseases; common risk factor; kernel density function; health risk curve

## 1. Introduction

For a long time, the diagnosis and treatment of diseases have been critical aspects of medical development. Many scholars have used data mining techniques on medical data to analyze the relationships between disorders and the real causes of those disorders [1–4]. Additionally, intelligent medical systems have been developed to help doctors diagnose illnesses [5–7]. However, abnormalities in physiological indicators may be a gauge of not only one disease but of multiple diseases. Therefore, in recent years, determining the common risk factors and developing a predictor model for multiple diseases have become more important. Chang et al. [8] proposed a two-stage analysis procedure that used data mining techniques and mathematical approaches to determine the common risk factors (such as systolic blood pressure (SBP), triglycerides (TGs), uric acid, glutamate pyruvate transaminase, and gender) and predictive models for hypertension and hyperlipidemia. Medical decision systems, based on analyzing risk factors and predicting the functions of diseases, can help patients understand the risks of developing diseases and efficiently provide diagnostic references for medical personnel. In the current era of medicine, preventive medicine has gradually become more accepted. The motivation of this study was to propose a new health risk index to assess the health status of people from the perspective of preventive medicine. The rapid development of medical devices has made it easy to

obtain information about physiological indicators. In the past, when all physiological indicators were within the standard range, the probability of illness was small and the patient was likely healthy. However, such information does not help the concept of preventive medicine. Therefore, a novel index must be developed to assess health status when there are changes in risk factors in healthy people.

Certain health care practices have also recently become more accepted; for example, people have started to use exercise, meditation, and diet to control risk factors. Whelton et al. [9] studied the effect of aerobic exercise on blood pressure and found that SBP and diastolic blood pressure (DBP) decreased by an average of 3.84 and 2.58 mmHg, respectively, over a period of aerobic training in 2419 subjects aged between 21 and 79 years old. Additionally, Fagard [10] designed a one-year experiment to observe the effects of exercise and diet on blood pressure; SBP and DBP decreased by an average of 3.4 and 2.4 mmHg, respectively, in 68 subjects after the implementation of an exercise regimen. Fagard [10] also found that exercise combined with dietary control resulted in a more significant effect than only exercise or only dietary control. Stewart et al. [11] designed an exercise plan for 115 subjects aged between 55 and 75 years old who were divided into two groups: a control group and a group that exercised three times a week. The results demonstrated that exercise could improve physical fitness and decrease the risk of many physiological factors related to cardiovascular diseases and diabetes. Based on the above discussion, exercise maintenance can reduce the values of several risk factors.

However, there has been little research on a method to evaluate the health status of a healthy individual when their risk factors change even after exercise maintenance. For example, hypertension patients are defined by an SBP of >140 mmHg or a DBP of >90 mmHg. When the SBP ranges between 120 and 135 mmHg, which is within the normal range, the resulting health effects may differ in different individuals. Vasan et al. found that in 9845 subjects with normal blood pressure, only 5.3% became hypertensive after four years; however, 17.6% of the subjects who developed hypertension originally had an SBP between 120 and 129 mmHg and a DBP between 80 and 84 mmHg [12]. The American Diabetes Association defines diabetes as a fasting blood glucose level of >126 mg/dL. The normal fasting blood glucose level for non-diabetics should range between 70 and 100 mg/dL. Although a fasting blood glucose level of <100 mg/dL is normal, a fasting blood glucose level between 100 and 125 mg/dL is considered pre-diabetic. Nichols et al. found that 8.1% of patients with fasting glucose levels between 100 and 109 mg/dL and 24.3% of those with levels between 110 and 125 mg/dL developed diabetes after an average of 41.4 months [13]. Therefore, risk factor values that are closer to the threshold will increase the risk of developing the respective disease. From the perspective of preventive medicine, understanding the degenerating state of healthy individuals will help prevent future disease.

Maintainability is widely used in industrial applications. Maintainability can be used to assess the life of a machinery or production system. The physical structures of the human body contain similar production systems; thus, an approach transferring the concepts of maintainability to human health assessment is the focus of this study. In practical applications, information on risk factors (such as blood pressure, heart rate, Electroencephalography (EEG), and Electromyography (EMG)) can be collected quickly with the widespread use of home medical equipment. Therefore, in this study, a kernel density technique was used to develop a health risk index based on risk factors from human-related data. In practice, changes in health risk indicators can be used to establish an early warning mechanism. When the health risk indicator is gradually declining, the patient will still be considered healthy; however, if they fail to control the risk factor, they will likely develop the disease. Thus, a health risk index can estimate the health status of people and help them better understand their health conditions. Additionally, manufacturers should apply new techniques [14–16] to develop rapid physiological devices or sensors to measure most risk factors. For example, Zhang et al. presents state-of-the-art research progress on cardiovascular health informatics and focuses on three major challenges: unobtrusive and wearable multi-parameter sensors, fast multimodal imaging technologies, and novel multi-scale information fusion heart models [17].

In this paper, the health risk index was proposed to evaluate a person's level of health. Figure 1 displays the framework of this study. The differences from previous methods are primarily a result of different study purposes. Previous research has developed disease prediction models through algorithmic designs by collecting information about specific diseases in healthy and unhealthy populations. Those algorithms use feature selection techniques to identify risks and establish a classification model by collecting data on risk factors. Their purpose is to determine whether a person is healthy or sick by measuring risk factors. Conversely, this study focuses on the assessment of healthy people. The similarity lies in the determination of risk factors, whereas the difference lies in the development of health risk indicators based on the data of healthy people. The purpose of this study was to assess health outcomes, particularly in high-risk groups. When health risk indicators decrease, individuals may develop diseases or otherwise become unhealthy. At this point, risk factors must be controlled to avoid disease. The results of this study can be used to jointly develop AI in healthcare through cloud computing to evaluate the trends and changes of users' health. This information is important for disease prevention.

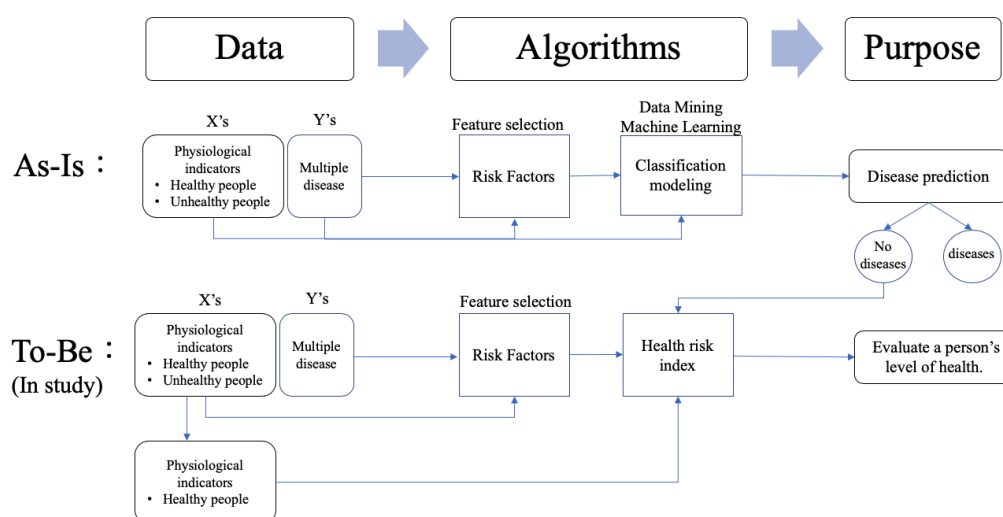


Figure 1. The research framework

## 2. Materials and Methods

This study proposed a novel three-stage analysis procedure involving the feature selection method, the kernel density estimation method, and mathematical approaches to calculate the health risk index of healthy people. Stage 1 adopted the findings from our previous study [8]. The procedure used six classification techniques to individually screen for the key risk factors for multiple diseases. Based on the results of previous research [8], we used five common risk factors: fasting plasma glucose (FPG), total cholesterol (T-CHO), TGs, SBP, and DBP, to determine the risks for hypertension, hyperlipidemia, and hyperglycemia. After identifying these risk factors, we compared the risk factors for each disease to determine the common risk factors for multiple diseases. Stage 2 used the kernel density estimation method to fit density curves for the common risk factors individually. Finally, stage 3 calculated the health risk index based on the kernel density function of the common risk factors. The two primary methods of kernel density estimation and the calculation of the health risk index are described below. The proposed methodology is explained below using data on 6496 subjects (3104 males and 3392 females) with Triple H disease from the National Health Insurance Research Database in Taiwan and the research results of Stewart et al. [11], who conducted a six-week exercise intervention. The reason for applying Stewart et al. [11] research data is that the five risk factors proposed are the same as the five common risk factors in this study. Also, the effectiveness of preventive maintenance was evaluated using a health risk curve.

### 2.1. Kernel Density Estimation Method

The kernel density estimation approach was first described by Rosenblatt [17] and Parzen [18]. It is a nonparametric statistical method used to estimate an unknown probability distribution. The method does not require a priori knowledge or make any additional assumptions regarding data distribution. In practice, it is often assumed that the values of the risk factor follow a normal distribution; however, this assumption lacks strong supporting evidence. Therefore, in this study, we adopted the kernel density method to estimate the risk factor distribution.

If  $x_1, x_2, \dots, x_n$  are independent and identically distributed unknown observations, the probability density function  $\hat{f}_h$  can be estimated by the kernel density function as follows:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where  $K(\cdot)$  is a kernel function that was symmetric and integrated into one. The  $h$  variable is the bandwidth to determine the degree of smoothness of the kernel function. This study evaluated six types of common kernel functions, including Gaussian, Epanechnikov, Triangular, Uniform, Bright, and Cosine, to estimate the kernel density values for all the common risk factors. Table 1 shows the six common types of kernel functions.

**Table 1.** The common types of kernel function.

Types	Kernel Function
Uniform	$K(u) = \frac{1}{2}$ for $ u  \leq 1$
Triangular	$K(u) = (1 -  u )$ for $ u  \leq 1$
Biweight	$K(u) = \frac{15}{16}(1 - u^2)^2$ for $ u  \leq 1$
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2)$ for $ u  \leq 1$
Gaussian	$K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$
Cosine	$K(u) = \frac{\pi}{4}\text{Cos}\left(\frac{\pi}{2}u\right)$ for $ u  \leq 1$

When using the kernel density estimation method, we must choose the kernel function and set the bandwidth. The choice of kernel function was adopted through the six common types in Table 1. To select the kernel function, all values of the risk factors were initially used to calculate the probability density value as the real density value. Then, the six common types of kernel functions in Table 1 were used to estimate the kernel density function for the risk factor as the estimated density value. Finally, a suitable kernel function was chosen using the minimum difference between the sum of the real density value and the estimated density value. Therefore, each density function of the risk factors might be fitted with different kernel functions.

The bandwidth of the kernel function is an important parameter that has a strong influence on the resulting estimate. A narrow bandwidth would allow more over-fitting of the data. Conversely, an overly wide bandwidth would not have an appropriate data fit. Figure 2 shows the differences from 100 standard normally distributed datapoints. When the bandwidth was set to 0.05 or 2, there was a large difference from the original density distribution; conversely, if the bandwidth was set to 0.337, there was an excellent fit.

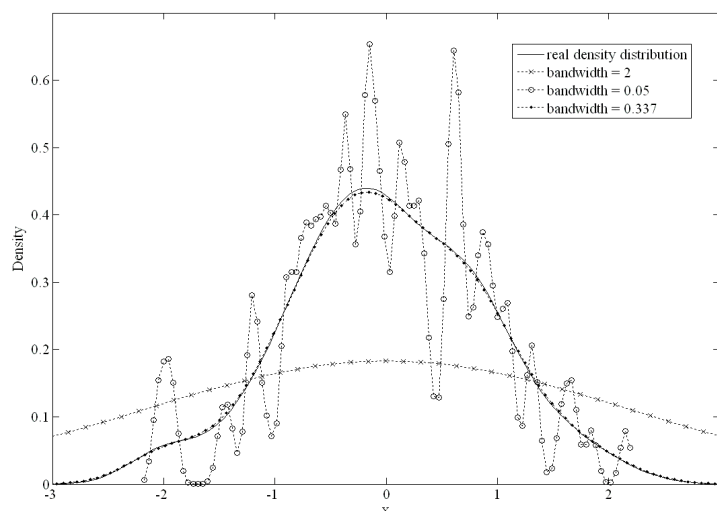


Figure 2. The differences with different bandwidth setting.

The mean integrated squared error (*MISE*) was used to choose the best bandwidth. The *MISE* can be calculated as follows:

$$MISE(h) = \int Bias^2(\hat{f}_h(x))dx + \int Var(\hat{f}_h(x))dx \tag{2}$$

The *MISE* was separated into two parts including  $Bias(\hat{f}_h(x)) = E[\hat{f}_h(x)] - f(x)$  and  $Var(\hat{f}_h(x)) = E[\hat{f}_h^2(x)] - E^2[\hat{f}_h(x)]$ .  $\hat{f}_h(x)$  was the estimated density value using the kernel function, and  $f(x)$  was the unknown true density value.  $\int Bias^2(\hat{f}_h(x))dx$  using the Taylor expansion method was derived as  $1/4[h^4\mu_2^2R(f)]$ ,  $\mu_2 = \int z^2K(z)dz$  and  $R(f) = \int f''(x)^2dx$ .  $n$  and  $h$  were the sizes of the data and bandwidth, respectively. If  $n$  was large and  $h$  was small, the variance of  $\hat{f}_h(x)$  can be derived as  $(1/nh) \int K^2(z)dz$ . Therefore, an approximate mean integrated squared error (*AMISE*) was calculated as follows:

$$AMISE = \frac{1}{4}h^4\mu_2^2R(f) + \frac{1}{nh} \int K^2(z)dz \tag{3}$$

We estimated the optimal bandwidth by minimizing *AMISE* regarding  $h$  by the first derivative. The optimal bandwidth  $h$  was:

$$\hat{h}_{AMISE} = \left[ \frac{\int K^2(z)dz}{\mu_2^2R(f)} \right]^{\frac{1}{5}} n^{-\frac{1}{5}} \tag{4}$$

In practice, the above formula has an infinite loop problem when calculating both  $\int K^2(z)dz$  and an unknown function  $f''(x)$ . Liu et al. compared the accuracies of nine types of bandwidth selection methods and found that no particular method performed better for all problems [19]. Therefore, in this study, we used the NDR0 method, which was suggested by Liu et al. [19]. The NDR0 method has the advantage of being easily calculated using the standard deviation ( $\hat{\sigma}$ ) and inter-quartile range of the dataset. The bandwidth can be calculated using the NDR0 as follows:

$$\hat{h} = 0.9\min\left(\hat{\sigma}, \frac{IQR}{1.34}\right)n^{-\frac{1}{5}} \tag{5}$$

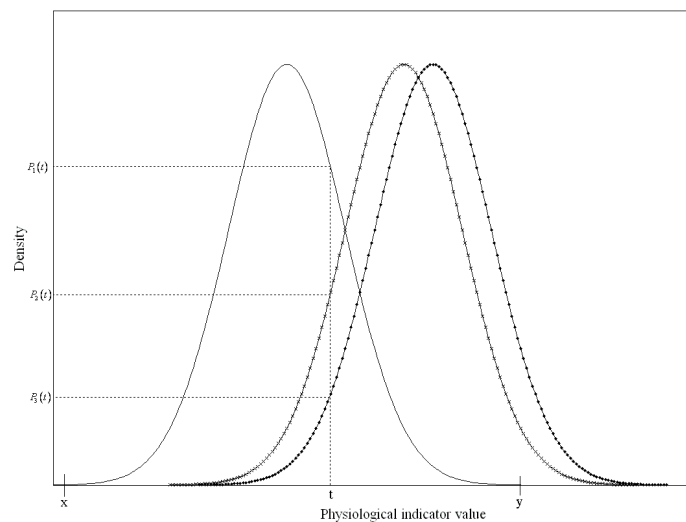
From the above procedure, we can determine the probability density function of the risk factors using the kernel density estimation approach. Next, the health risk index was proposed to estimate the health status of healthy people.

### 2.2. Health Risk Index Calculation

The  $R(t)$  was defined as a function of human health evaluation; it describes the  $t$  value of the risk factor once it has reached a certain value for a physiological state that was still healthy. In other words,  $R(t)$  is the health risk at value  $t$  for a particular risk factor. With an increase in the risk factor value, the disease or health risk will also increase. We assumed that  $n$  of multiple diseases were studied, and there was one normal state (healthy people) and  $(2^n - 1)$  combinations of people who suffered from different diseases. The  $f_{1,i}$  variable is representative of healthy individuals with the  $i^{\text{th}}$  risk factor in the probability density function. The  $f_{2,i}, f_{3,i}, \dots, f_{2^n,i}$  variables represent the probability density function of the different combinations of diseases with the  $i^{\text{th}}$  risk factor. The probability density function used the kernel density estimation approach. In this study, the health risk index  $R_i(t)$  was defined for healthy people with the  $i^{\text{th}}$  risk factor and was calculated as follows:

$$R_i(t) = \begin{cases} \frac{f_{1,i}}{f_{1,i} + f_{2,i} + \dots + f_{2^n,i}} & , \text{for } x_i \leq t \leq y_i \\ 0 & , \text{others} \end{cases} \quad (6)$$

where  $t$  represents the value of the  $i^{\text{th}}$  risk factor, and the interval  $[x_i, y_i]$  represents the range of values in the  $i^{\text{th}}$  risk factors for healthy people. Figure 3 presents an example of three different status functions of a risk factor to describe the health risk index. The solid line represents the function of the risk factor values of healthy subjects, and the other two lines with signs represent groups that suffered from different diseases.



**Figure 3.** The relationship between probability density function and health risk index.

Since the  $t$  value of the risk factor only fell onto the  $x$  and  $y$ -axes for healthy subjects when it was between  $x$  and  $y$ ,  $R(t)$  was the ratio of the probability density of the normal group  $P_1(t)$  to all of the probability densities of the three status functions. Therefore,  $R(t)$  can be simplified as:

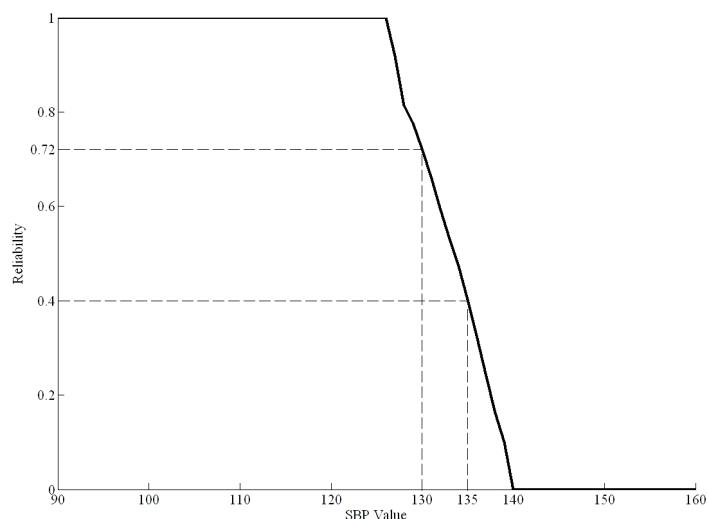
$$R(t) = \frac{P_1(t)}{p_1(t) + p_2(t) + p_3(t)}, \text{ for } x \leq t \leq y. \quad (7)$$

Each probability density value was a positive number less than 1. Therefore,  $R(t)$  was between 0 and 1.

If any of the risk factors were over the threshold values, we determined that the person had the disease(s). Therefore, the health risk index  $R_{hri}(t)$  for a healthy person can be presented as follows:

$$R_{hri}(t) = \prod_{i=1}^n R_i(t) \quad (8)$$

We can plot the curve of the health risk index from  $R_{hri}(t)$  for all  $t$  values of the risk factors. According to the health risk index curve, we can then evaluate the effect on health risks when risk factors change after a period of maintenance activities (such as exercise or diet control). Figure 4 shows an example of the health risk curve for SBP from 90 to 160 mmHg based on real data.



**Figure 4.** The health risk curve of systolic blood pressure (SBP).

In Figure 4, when the SBP is greater than 127 mmHg, the health risk curve begins to decline, indicating a gradual decline in health. When the SBP is greater than 140 mmHg, the health risk index is zero, indicating that the individual has the disease. For example, if a healthy person had an SBP of 135 mmHg before dietary control, the health risk index  $R_{hri}(t)$  of SBP would be 0.4. Moreover, if their SBP decreased to 130 mmHg after a period of dietary control, the new health risk index  $R_{hri}(t)$  of SBP would be 0.72. Therefore, the difference value between 0.4 and 0.72 (0.32) is the effect of the dietary control.

### 3. Results and Discussion

Hypertension, hyperlipidemia, and hyperglycemia are three common diseases that are associated with metabolic syndrome and related to metabolic abnormalities. If a person has these diseases, the risk of developing other chronic diseases increases as well. A dataset from the National Health Insurance Research Database in Taiwan was used to elucidate the proposed analytical procedures. The dataset included 6496 subjects (3104 males and 3392 females) aged over 15 years old and 17 physiological indicators.

Next, the analysis procedure established the appropriate kernel function for the common risk factors in every physiological state of hypertension, hyperlipidemia, and hyperglycemia. Tables 2–6 show the risk factors of the kernel density functions that were selected by the proposed estimation approaches used in this study. An example in Table 2 describes the results and indicates the outcome of the kernel density approach for FPG in the different physiological states (hypertension, hyperlipidemia, and hyperglycemia). In the normal state, the minimum value of the estimated six kernel density approaches was 0.0725. Therefore, the Gaussian type was selected as a suitable kernel function for FPG in the normal state. Similarly, the minimum value was 1.0622 for hyperlipidemia + hyperglycemia and 1.1331 for hypertension + hyperlipidemia + hyperglycemia. The Gaussian type was also the suitable kernel function for FPG in hyperlipidemia + hyperglycemia and hypertension + hyperlipidemia + hyperglycemia. The Triangular type was the appropriate kernel function for FPG for hypertension, hyperlipidemia, hyperglycemia, hypertension + hyperlipidemia, and hypertension + hyperglycemia, with the minimum values shown in Table 2.

**Table 2.** The selection of Kernel density function for the fasting plasma glucose (FPG).

FPG	Kernel Density Function					
	Gaussian	Epanechnikov	Uniform	Triangular	Biweight	Cosine
Normal	0.0725	0.0761	0.0781	0.0731	0.0753	0.0749
Hypertension	0.2387	0.2430	0.2508	0.2376	0.2425	0.2419
Hyperlipidemia	0.1242	0.1304	0.1374	0.1242	0.1286	0.1279
Hyperglycemia	0.9441	0.9485	0.9492	0.9428	0.9473	0.9468
(Hypertension & Hyperlipidemia)	0.2001	0.2052	0.2127	0.1995	0.2039	0.2033
(Hypertension & Hyperglycemia)	1.4843	1.4868	1.4935	1.4826	1.4863	1.4860
(Hyperlipidemia & Hyperglycemia)	1.0622	1.0658	1.0668	1.0624	1.0647	1.0643
(Hypertension & Hyperlipidemia & Hyperglycemia)	1.1331	1.1383	1.1439	1.1332	1.1367	1.1361

**Table 3.** The selection of Kernel density function for the total cholesterol (T-CHO).

T-CHO	Kernel Density Function					
	Gaussian	Epanechnikov	Uniform	Triangular	Biweight	Cosine
Normal	0.1184	0.1228	0.1281	0.1198	0.1214	0.1209
Hypertension	0.3555	0.3649	0.3781	0.3557	0.3613	0.3601
Hyperlipidemia	0.3247	0.3368	0.3404	0.3290	0.3328	0.3315
Hyperglycemia	0.8521	0.8631	0.8690	0.8532	0.8600	0.8587
(Hypertension & Hyperlipidemia)	0.4810	0.4864	0.4985	0.4796	0.4847	0.4841
(Hypertension & Hyperglycemia)	1.0488	1.0525	1.0584	1.0470	1.0515	1.0510
(Hyperlipidemia & Hyperglycemia)	0.9328	0.9400	0.9488	0.9331	0.9376	0.9368
(Hypertension & Hyperlipidemia & Hyperglycemia)	1.1214	1.1237	1.1235	1.1197	1.1230	1.1227

**Table 4.** The selection of Kernel density function for the triglyceride (TG).

TG	Kernel Density Function					
	Gaussian	Epanechnikov	Uniform	Triangular	Biweight	Cosine
Normal	0.1542	0.1569	0.1605	0.1542	0.1559	0.1555
Hypertension	0.4502	0.4506	0.4492	0.4490	0.4506	0.4506
Hyperlipidemia	0.3936	0.3973	0.4019	0.3945	0.3963	0.3959
Hyperglycemia	1.1723	1.1753	1.1776	1.1699	1.1745	1.1742
(Hypertension & Hyperlipidemia)	0.7481	0.7513	0.7497	0.7491	0.7504	0.7500
(Hypertension & Hyperglycemia)	1.4460	1.4466	1.4408	1.4404	1.4470	1.4470
(Hyperlipidemia & Hyperglycemia)	1.4483	1.4489	1.4411	1.4481	1.4487	1.4486
(Hypertension & Hyperlipidemia & Hyperglycemia)	1.5712	1.5751	1.5875	1.5712	1.5737	1.5732

**Table 5.** The selection of Kernel density function for the systolic blood pressure (SBP).

SBP	Kernel Density Function					
	Gaussian	Epanechnikov	Uniform	Triangular	Biweight	Cosine
Normal	0.1858	0.2023	0.2047	0.1914	0.1964	0.1947
Hypertension	0.4879	0.5139	0.5477	0.4952	0.5051	0.5021
Hyperlipidemia	0.2450	0.2562	0.2753	0.2474	0.2528	0.2519
Hyperglycemia	0.6924	0.7011	0.7258	0.6821	0.6983	0.6970
(Hypertension & Hyperlipidemia)	0.4440	0.4853	0.5307	0.4584	0.4709	0.4661
(Hypertension & Hyperglycemia)	1.1052	1.1400	1.1756	1.1091	1.1275	1.1232
(Hyperlipidemia & Hyperglycemia)	0.5720	0.5863	0.5930	0.5712	0.5811	0.5795
(Hypertension & Hyperlipidemia & Hyperglycemia)	0.8430	0.8601	0.8902	0.8378	0.8545	0.8523



**Table 6.** The selection of Kernel density function for diastolic blood pressure (DBP).

DBP	Kernel Density Function					
	Gaussian	Epanechnikov	Uniform	Triangular	Biweight	Cosine
Normal	0.2350	0.2585	0.2459	0.2463	0.2487	0.2459
Hypertension	0.5148	0.5454	0.5393	0.5201	0.5339	0.5295
Hyperlipidemia	0.2434	0.2656	0.2722	0.2501	0.2578	0.2557
Hyperglycemia	0.5583	0.5688	0.5843	0.5502	0.5657	0.5647
(Hypertension & Hyperlipidemia)	0.4065	0.4329	0.5135	0.4163	0.4276	0.4242
(Hypertension & Hyperglycemia)	0.9539	0.9985	1.0707	0.9558	0.9801	0.9745
(Hyperlipidemia & Hyperglycemia)	0.3873	0.4009	0.4333	0.3878	0.3962	0.3948
(Hypertension & Hyperlipidemia & Hyperglycemia)	0.6384	0.6800	0.7062	0.6471	0.6642	0.6591

Based on the results in Tables 2–7 summarizes all kernel density functions for the five common risk factors in every physiological state.

**Table 7.** The suitable kernel function for five risk factors.

Kernel Function	Risk Factors				
	FPG	T-CHO	TG	SBP	DBP
Normal	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Hypertension	Triangular	Gaussian	Triangular	Gaussian	Gaussian
Hyperlipidemia	Triangular	Gaussian	Gaussian	Gaussian	Gaussian
Hyperglycemia	Triangular	Gaussian	Triangular	Triangular	Triangular
(Hypertension & Hyperlipidemia)	Triangular	Triangular	Gaussian	Gaussian	Gaussian
(Hypertension & Hyperglycemia)	Triangular	Triangular	Triangular	Gaussian	Gaussian
(Hyperlipidemia & Hyperglycemia)	Gaussian	Gaussian	Rectangular	Triangular	Gaussian
(Hypertension & Hyperlipidemia & Hyperglycemia)	Gaussian	Triangular	Gaussian	Triangular	Gaussian

Based on the data in Table 7 and the proposed formula, the health risk curve of the five risk factors in the normal state was calculated and plotted (Figure 5). Figure 5a–d displays the health risk curves for FPG, T-CHO, TG, SBP, and DBP. The health risk curves of the risk factors display monotonic decreasing curves, as shown in Figure 4. When the FPG reached 126 mg/dL, T-CHO or TG reached 200 mg/dL, SBP reached 140 mmHg, or DBP reached 90 mmHg, the health risk value became 0. When any risk factor value remained over the threshold, the probability that the subject was in the normal physical state was 0. The health risk index of the physiological system could be determined by multiplying the health risk values of the risk factors. Table 8 shows the results of ten subjects who were selected from the dataset.

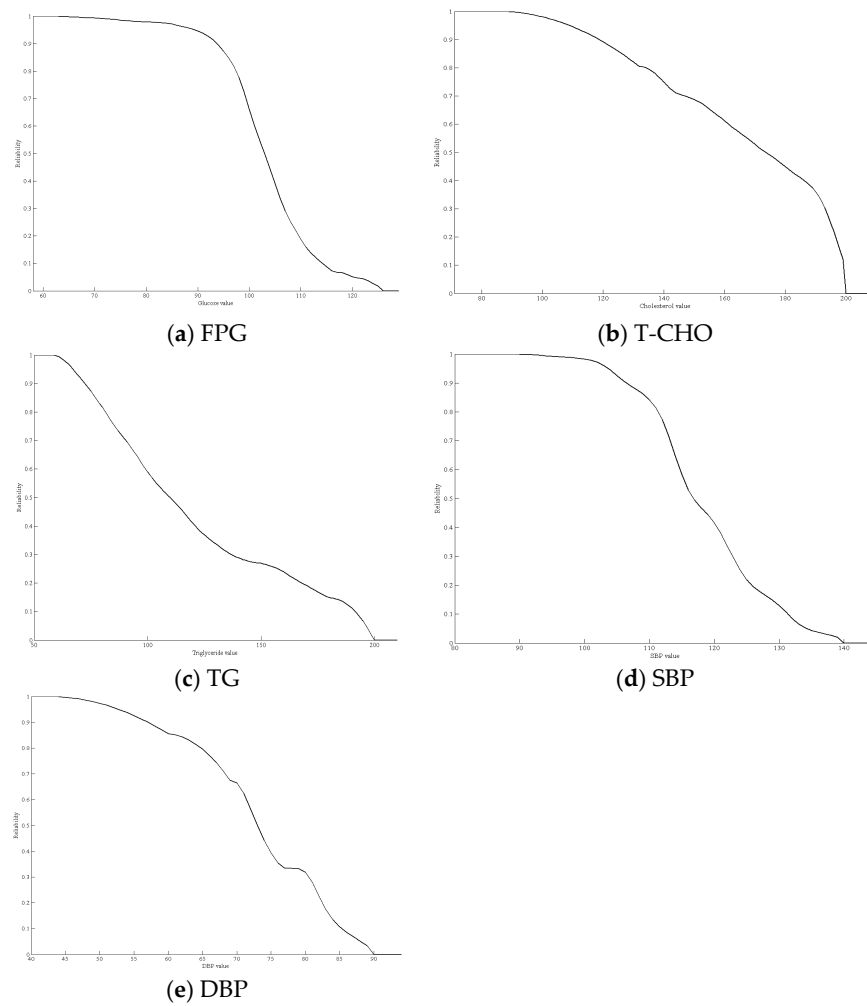


Figure 5. Health risk curves for the five risk factors.

Table 8. The result of ten subjects for Health risk index.

Subjects	FPG	T-CHO	TG	SBP	DBP	Physiological State	Health Risk Index
1	89 (0.9527)	177 (0.4751)	153 (0.2632)	141 (0.0000)	90 (0.0000)	Hypertension	0
2	88 (0.9579)	185 (0.4100)	109 (0.5070)	124 (0.2521)	92 (0.0000)	Hypertension	0
3	86 (0.9672)	119 (0.8985)	56 (1.0000)	97 (0.9899)	60 (0.8552)	Normal	0.7357
4	238 (0.0000)	190 (0.3606)	103 (0.5591)	140 (0.0000)	69 (0.6752)	(Hypertension & Hyperglycemia)	0
5	80 (0.9793)	143 (0.7189)	53 (1.0000)	133 (0.0663)	67 (0.7432)	Normal	0.0347
6	105 (0.3917)	167 (0.5543)	99 (0.6012)	92 (0.9976)	60 (0.8552)	Normal	0.1114
7	79 (0.9800)	169 (0.5375)	117 (0.4378)	101 (0.9785)	73 (0.4997)	Normal	0.1128
8	97 (0.8187)	234 (0.0000)	263 (0.0000)	112 (0.7724)	80 (0.3183)	Hyperlipidemia	0
9	81 (0.9785)	167 (0.5543)	87 (0.7346)	109 (0.8627)	61 (0.8512)	Normal	0.2926
10	89 (0.9527)	152 (0.6766)	100 (0.5901)	110 (0.8425)	71 (0.6256)	Normal	0.2005

Ten subjects were randomly recruited to measure the risk factor data. The brackets in Table 8 indicate the health risk index values. For example, the FPG value for subject 1 was 89 (0.9527); this indicated that the health risk index of an FPG level of 89 mg/dL was 0.9527. The health risk index of the physiological system of several subjects was 0 because some of the risk factor values were over the clinical thresholds, such as in Subjects 1, 2, 4, and 8. Subject 4 suffered from hypertension and hyperglycemia. The six healthy people were Subjects 3, 5, 9, 10, 7, and 6, arranged in descending order according to their health risk indexes. Although the six subjects had normal health statuses, the reliability values of the physiological systems differed. When applying the health index, the study of Stewart et al. [11] was used as a reference. Stewart et al. [11] simulated the effect of five risk factors after implementing an exercise regimen. After a six-week exercise routine, FPG, T-CHO, TG, SBP, and DBP levels decreased an average of 0.2 mg/dL, 5.2 mg/dL, 13.4 mg/dL, 5.3 mmHg, and 3.7 mmHg, respectively. We determined the change in the risk factors by calculating the health risk index. The results showed that subjects experienced a 7.29% reduction in their health risk index for hypertension, hyperlipidemia, and hyperglycemia.

The dataset on hypertension, hyperlipidemia, and hyperglycemia (Triple H) from the National Health Insurance Research Database in Taiwan and the research results of Stewart et al. [11] were used to explain the proposed methodology and application. Previous research has focused on disease prediction and medical treatment after disease identification. If judged as healthy, there is no corresponding early warning mechanism. The difference between the proposed method and the previous method is that this study develops an evaluation health risk index from the perspective of probability reliability. The estimation of the optimal probability distribution is established using the kernel density approach rather than using a normal probability function. The research results can be applied to solve the problem of health warning. When people are considered healthy, health risk indicators can be further calculated to understand their level of health. This benefits the individuals because risk factors can be controlled based on fitness by implementing strategies to increase exercise or control diet. Furthermore, the health risk index can be used to evaluate the effectiveness of the strategy and adjust the strategy. This will concretely improve the effective use of resources in the medical/health field.

#### 4. Conclusions

Based on the analysis of the results, we can make several conclusions regarding research in this area. The human body is a complex system, and a single-disease study is insufficient to assess multiple diseases. Most previous studies that conducted medical data analyses focused on risk factor selection and creating prediction and control models of diseases. These studies were able to monitor risk factors and estimate the risk of developing various diseases. However, the human body may not necessarily be completely healthy even when risk factors are all within normal ranges. Therefore, our study focused on assessing the degree of health. In this paper, common risk factors for multiple diseases were used to estimate the optimal probability density function, and a novel health risk index was proposed to evaluate the health status of healthy people.

In practice, a health risk curve can be used to elucidate the relationship between risk factors and health risk. Prevention is often more desirable than curing diseases, and proper maintenance (such as participating in sports) may reduce the likelihood of further health decline. Furthermore, the effectiveness of preventive maintenance can be assessed through a health risk curve to determine appropriate adjustments to the maintenance strategy.

**Author Contributions:** C.-C.W., C.-D.C. and B.C.J. designed the study. C.-D.C. was responsible for data collection and analysis. C.-D.C., C.-C.W. and B.C.J. reviewed relevant literature and interpreted the acquired data. C.-C.W. and C.-D.C. drafted the manuscript. Both authors have read and approved the final manuscript.

**Funding:** This research was funded by Ministry of Science and Technology for supporting this research with ID MOST 101-2221-E-131-030-MY3.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ghorbani, R.; Ghousi, R. Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. *Int. J. Data Netw. Sci.* **2019**, *3*, 47–70. [[CrossRef](#)]
2. Su, C.T.; Wang, P.C.; Chen, Y.C.; Chen, L.F. Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients. *J. Med. Syst.* **2012**, *36*, 2387–2399. [[CrossRef](#)] [[PubMed](#)]
3. Deng, X.; Yu, T.; Hu, A. Predicting the Risk for Hospital-Acquired Pressure Ulcers in Critical Care Patients. *Crit. Care Nur.* **2017**, *37*, e1–e11. [[CrossRef](#)] [[PubMed](#)]
4. Wong, B.; Wong, B.; Ho, G.T.S.; Tsui, E.; Tsui, E. Development of an intelligent e-healthcare system for the domestic care industry. *Ind. Manag. Data Syst.* **2017**, *117*, 1426–1445. [[CrossRef](#)]
5. Bath, P.A. Data mining in health and medical information. *Annu. Rev. Inf. Sci. Technol.* **2004**, *38*, 331–369. [[CrossRef](#)]
6. Liberatore, M.J.; Nydick, R.L. The analytic hierarchy process in medical and health care decision making: A literature review. *Eur. J. Oper. Res.* **2008**, *189*, 194–207. [[CrossRef](#)]
7. Crothers, K.; Kross, E.K.; Reisch, L.M.; Shahrir, S.; Slatore, C.; Zeliadt, S.B.; Triplette, M.; Meza, R.; Elmore, J.G. Patients' attitudes regarding lung cancer screening and decision aids. A survey and focus group study. *Ann. Am. Thorac. Soc.* **2016**, *13*, 1992–2001. [[CrossRef](#)] [[PubMed](#)]
8. Chang, C.D.; Wang, C.C.; Jiang, B.C. Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors. *Expert Syst. Appl.* **2011**, *38*, 5507–5513. [[CrossRef](#)]
9. Whelton, S.P.; Chin, A.; Xin, X.; He, J. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. *Ann. Int. Med.* **2002**, *136*, 493–503. [[CrossRef](#)] [[PubMed](#)]
10. Fagard, R.H. Effects of exercise, diet and their combination on blood pressure. *J. Hum. Hypertens.* **2005**, *19*, S20. [[CrossRef](#)] [[PubMed](#)]
11. Stewart, K.J.; Bacher, A.C.; Turner, K.; Lim, J.G.; Hees, P.S.; Shapiro, E.P.; Tayback, M.; Ouyang, P. Exercise and risk factors associated with metabolic syndrome in older adults. *Am. J. Prev. Med.* **2005**, *28*, 9–18. [[CrossRef](#)] [[PubMed](#)]
12. Vasan, R.S.; Larson, M.G.; Leip, E.P.; Kannel, W.B.; Levy, D. Assessment of frequency of progression to hypertension in non-hypertensive participants in the Framingham Heart Study: A cohort study. *Lancet* **2001**, *358*, 1682–1686. [[CrossRef](#)]
13. Nichols, G.A.; Hillier, T.A.; Brown, J.B. Progression from newly acquired impaired fasting glucose to type 2 diabetes. *Diabetes Care* **2007**, *30*, e80. [[CrossRef](#)]
14. Pan, J.S.; Kong, L.; Sung, T.W.; Tsai, P.W.; Snášel, V.  $\alpha$ -Fraction First Strategy for Hierarchical Model in Wireless Sensor Networks. *J. Int. Technol.* **2018**, *19*, 1717–1726.
15. Wu, W.; Pirbhulal, S.; Zhang, H.; Mukhopadhyay, S.C. Quantitative assessment for self-tracking of acute stress based on triangulation principle in a wearable sensor system. *IEEE J. Biomed. Health Inf.* **2019**, *23*, 703–713. [[CrossRef](#)] [[PubMed](#)]
16. Bradley, C.; Bowery, A.; Britten, R.; Budelmann, V.; Camara, O.; Christie, R.; Cookson, A.; Frangi, A.F.; Gamage, T.B.; Heidlauf, T.; et al. OpenCMISS: A multi-physics & multi-scale computational infrastructure for the VPH/Physiome project. *Prog. Biophys. Mol. Biol.* **2011**, *107*, 32–47. [[PubMed](#)]
17. Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **1956**, *27*, 832–837. [[CrossRef](#)]
18. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [[CrossRef](#)]
19. Liu, B.; Yang, Y.; Webb, G.I.; Boughton, J. A comparative study of bandwidth choice in kernel density estimation for naive Bayesian classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 302–313.

