Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/020301

Sensitivity of Health Related Indices is a Non-Decreasing Function of their Partitions

Fragiskos Bersimis[1], *Demosthenes Panagiotakos* [2,∗] *and Malvina Vamvakari*[1]

[1] Department of Informatics and Telematics, Harokopio University, 70 Eleftheriou Venizelou Str., Athens, Greece
[2] Department of Nutrition Science-Dietetics, Harokopio University, 70 Eleftheriou Venizelou Str., Athens, Greece

Received: 12 Mar. 2013, Revised: 24 Jun. 2013, Accepted: 25 Jun. 2013
Published online: 1 Nov. 2013

Abstract: This work aims to theoretically evaluate whether the number k of partitions of a discrete variable X affects the sensitivity $S_e(X)$ of a binary health outcome Y. The distribution of variable X was either unknown or the uniform. Thus, two discrete random variables X^k and X^{k+1} with k and $k+1$ partitions, respectively, were considered. In addition, a random variable Y that indicates the actual health status of an individual was also considered. The case of the composite index T_m^k which is formed by the sum of m variables X_j^k, $j = 1,2,\ldots,m$ either when the distribution of each variable X_j^k is unknown or the uniform was also investigated. This work suggests that the sensitivity of an index is a non-decreasing function of the number of partitions, under certain conditions.

Keywords: Composite health related index, sensitivity, non-decreasing function

1 Introduction

Health related indices have long been used in biomedical research [1,2]. Health related indices are composite tools, based on either discrete or continuous variables. This type of indices aims to measure a variety of clinical conditions, behaviors, attitudes and beliefs, which are difficult and even impossible to be measured quantitatively and directly (e.g. emotions, stress, depression, pain, diet quality, etc) [2,3]. In practice, a composite health related index T_m^k is created by the sum of m component variables X^k, each one with k partitions. Thus, a composite health related index T_m^k is given by the following formula:

$$T_m^k = \sum_{j=1}^{m} X_j^k \tag{1.1}$$

The m component variables reflect different m aspects of a person's clinical situation.

During the past years, indices have been extensively used in various health fields such as in psychometry in order to measure several conditions, like depression, anxiety, stress [4] as well as in cardiovascular prevention in order to measure diet quality and adequacy [5]. Based on these tools, individuals were classified as being related or non-related to the investigated characteristics of a specific disease. Moreover, health related indices have been associated with several health outcomes such as a diet scale being related to the likelihood of developing cardiovascular disease [6]. In addition, an index that measures stressful experiences has been related to sudden deaths, etc. As an example, MedDietScore [6] is a dietary composite index that measures the degree of adherence to the Mediterranean diet. The consumption of each of the 11 food groups (fruits, vegetables, meat, fish, dairy, etc) is calibrated depending on the frequency of consumption (e.g. never, rarely, 1-2 times per month etc). The calibration is either from 0 to 5 whether the food group is beneficial for health or from 5 to 0 whether the opposite is true (e.g. meat). The total value of this dietary index is the sum of the individual values received by the component food groups. A great value of this dietary index for a person means that, this person is close to adherence of the traditional Mediterranean diet pattern and thus runs less risk of cardiovascular disease. Thus, MedDietScore is a composite health related index with $m = 11$ components each of which has $k = 6$ partitions.

Despite the fact that health related indices have been used in many fields of biomedical research, there are several unresolved issues as regards their construction [2]. One of these issues is the optimal number of partitions of the index

∗ Corresponding author e-mail: dbpanag@hua.gr

needed, in order to increase, not only the information retrieved from the individual's characteristics, but also the diagnostic ability of the index as regards the health outcome that the index aims to evaluate. In a recent work by Kourlaba and Panagiotakos [7], it has been shown using simulated and empirical data that the number of partitions each component of a composite index has, is positively related to the diagnostic ability of the index as well as that the use of a continuous scale is the optimal choice to achieve the maximum diagnostic accuracy. However, from a theoretical viewpoint, simulations are not so robust to establish a methodology because not all possible cases, which may exist in a mathematical solution of a research hypothesis, can be covered. Therefore, simulation methods are not considered as a mathematical proof, although they are an analytical proof. As a result, such a proof is still missing in literature. Thus, the aim of this work was to evaluate the research hypothesis which claims that the sensitivity of an index increases when the partitions of the index also increase, based on a theoretical approach. The latter, has also been partially attempted, in a recent work by Maroulas and Panagiotakos [8]. In order to examine both the nature and the intensity of the relation between the number of partitions each component of a composite index has, as well as its diagnostic ability, a measure of the diagnostic accuracy of the index was necessary. The sensitivity function of a composite index is suitable to be used in order for the aforementioned research hypothesis to be investigated. It is known that sensitivity function itself is not sufficient enough to evaluate the diagnostic accuracy of an index as well as that the increase of sensitivity leads to the reduction of specificity and vice versa, by moving the diagnostic threshold. In this work the number of partitions is being examined according to the sensitivity value.

This paper has been organized into four sections. The first section presents the first statement of this work (Proposition 1), as well as its mathematical proof, in the case of a discrete variable X^k, which represents a health related index T_1^k with k partitions that follows an unknown distribution. The second section presents the special case where the discrete variable X^k follows the uniform distribution whereas the third section presents the extension of Proposition 1 in the case that the index T_m^k is developed by the sum of m variables X_j^k and their distribution is unknown (Proposition 3). Besides, the fourth section refers to the application of Proposition 3 and also presents specific examples of the sensitivity function's monotony of a composite index T_m^k which is the sum of m discrete variables, in the case that, variables X_j^k are distributed uniformly.

2 Problem Setup - One discrete Variable - General Results

In this section, the first statement of this work is presented, as well as its proof, in the case of a discrete variable X^k, which represents a health related index T_1^k. Thus, it is investigated, by the use of the sensitivity function, whether the number of partitions of the discrete variable X^k affects its predictive ability for a health outcome. Therefore, two discrete random variables X^k and X^{k+1} are considered with k and $k+1$ partitions, respectively.

Notation: As a k-partition of a set A is considered a set of k sets $U = \{U_i\}$, $i = 1,2,\ldots,k$ that are pairwise disjoint, not null subsets of A and their union is A [9].

The discrete variables X^k and X^{k+1} take values from the following sets:

$$R_{X^k} = \{1,2,\ldots,k\} \text{ and } R_{X^{k+1}} = \{1,2,\ldots,k,k+1\}.$$

A random bivariate variable Y is considered so as to follow the Bernoulli distribution [8]. Variable Y expresses the actual clinical status of a person as well as indicates the presence or not of a disease (healthy $Y = 0$ or patient $Y = 1$).

Sensitivity of a clinical test is the probability of a positive test result given the presence of the symptom [10]. Generally, sensitivity measures the proportion of actual positives (diseased) which are identified as such by the test result.

The sensitivity functions of variables X^k and X^{k+1}, in relation to Y, are defined according to the following conditional probabilities [10]:

$$S_e(X^k) = P(X^k > c_0 \mid Y = 1) \tag{2.1}$$

and

$$S_e(X^{k+1}) = P(X^{k+1} > c_0' \mid Y = 1) \tag{2.2}$$

where c_0 and c_0' are selected thresholds in an appropriate statistical method (e.g. by the Receiver Operating Characteristic Curve) and $c_0, c_0' \in \mathbb{R}$.

If π_i^k and π_i^{k+1} imply the conditional probabilities where the discrete random variables X^k and X^{k+1} take a fixed value i, for an individual of the diseased population $(Y = 1)$, then π_i^k and π_i^{k+1} are given from the following formulas:

$$\pi_i^k = P(X^k = i \mid Y = 1), i = 1,2,\ldots,k, \text{ and } k \in \mathbb{N}, k > 1,$$

and

$$\pi_i^{k+1} = P(X^{k+1} = i \mid Y = 1), i = 1,2,\ldots,k,k+1, \text{ and } k \in \mathbb{N}, k > 1.$$

Therefore, according to definitions 2.1 and 2.2 sensitivity's functions of variables X^k and X^{k+1} are given by the formulas 2.3 and 2.4:

$$S_e(X^k) = P(X^k \geq c \mid Y = 1) = \sum_{i=c}^{k} \pi_i^k \tag{2.3}$$

and

$$S_e(X^{k+1}) = P(X^{k+1} \geq c' \mid Y = 1) = \sum_{i=c'}^{k+1} \pi_i^{k+1} \tag{2.4}$$

where $c \in R_{X^k} = \{1, 2, \ldots, k\}$ is the lowest value of X^k greater than c_0 and $c' \in R_{X^{k+1}} = \{1, 2, \ldots, k, k+1\}$ is the lowest value of X^{k+1} greater than c'_0.

Proposition 1. *For any $k \in \mathbb{N}$, the sensitivity of index X^k, is a non-decreasing function of k, under a specific condition (condition 1).*

Proof. In order the above proposition to be proved, it is sufficient to be shown that for any $k \in \mathbb{N}$ the following inequality applies:

$$k < k+1 \Leftrightarrow S_e(X^k) \leq S_e(X^{k+1}) \tag{2.5}$$

According to definitions 2.3 and 2.4, relationship 2.5 becomes:

$$\sum_{i=c}^{k} \pi_i^k \leq \sum_{i=c'}^{k+1} \pi_i^{k+1} \tag{2.6}$$

A sequence of probabilities' differences $\alpha_i, i = 1, 2, \ldots, k$ is considered so as to satisfy the following equality:

$$\alpha_i = \pi_i^k - \pi_i^{k+1} \tag{2.7}$$

Condition 1: The sequence of probabilities' differences $\alpha_i, i = 1, 2, \ldots, k$ satisfies the following inequality

$$\sum_{i=1}^{c-1} \alpha_i \geq \sum_{i=c}^{c'-1} \pi_i^{k+1} \geq 0 \tag{c_1}$$

in the case that $c < c'$.

The sum on the left side of c_1 refers to the difference between the probabilities of false negative values of indices X^k and X^{k+1}. On the other hand, the sum on the right side of c_1 refers to the probabilities of false negative values of index X^{k+1}, between c and $c' - 1$. Practically, condition 1 declares that the probabilities' differences of false negative values, between the index with fewer partitions and the index with more partitions, cumulatively, should be greater than or equal to the probabilities' sum of false negative values of the index X^{k+1}, between c and $c' - 1$.

Therefore, the cases which are distinguished, according to the relative position between c and c' are the following:

–Case 1: $c = c' > 1$

According to relation 2.7, the sensitivity of variable X^k becomes:

$$\begin{aligned} S_e(X^k) &= \sum_{i=c}^{k} \pi_i^k = \sum_{i=c}^{k} \left(\pi_i^{k+1} + \alpha_i\right) = \sum_{i=c}^{k} \pi_i^{k+1} + \sum_{i=c}^{k} \alpha_i \overset{(c_1)}{\leq} \sum_{i=c}^{k} \pi_i^{k+1} + \sum_{i=1}^{c-1} \alpha_i + \sum_{i=c}^{k} \alpha_i \\ &= \sum_{i=c}^{k} \pi_i^{k+1} + \sum_{i=1}^{k} \alpha_i \overset{(l_1)}{=} \sum_{i=c}^{k} \pi_i^{k+1} + \pi_{k+1}^{k+1} = \sum_{i=c}^{k+1} \pi_i^{k+1} = S_e(X^{k+1}) \\ &\Leftrightarrow S_e(X^k) \leq S_e(X^{k+1}) \end{aligned}$$

And thus 2.5 is established. □

Lemma 1. *For the sequence of probabilities' differences α_i, $i = 1, 2, \ldots, k$ applies:*

$$\sum_{i=1}^{k} \alpha_i = \pi_{k+1}^{k+1} \tag{l_1}$$

*Proof.*By summing from $i=1$ to k both sides in relation 2.7 imply that:

$$\alpha_i=\pi_i^k-\pi_i^{k+1}\Leftrightarrow\sum_{i=1}^{k}\alpha_i=\sum_{i=1}^{k}\left(\pi_i^k-\pi_i^{k+1}\right)\Leftrightarrow\sum_{i=1}^{k}\alpha_i=\sum_{i=1}^{k}\pi_i^k-\sum_{i=1}^{k}\pi_i^{k+1}=1-\sum_{i=1}^{k}\pi_i^{k+1}=\sum_{i=1}^{k+1}\pi_i^{k+1}-\sum_{i=1}^{k}\pi_i^{k+1}=\pi_{k+1}^{k+1}$$

–**Case 2**: $c>c'>1$
According to relation 2.7, the sensitivity of variable X^k becomes:

$$S_e(X^k)=\sum_{i=c}^{k}\pi_i^k=\sum_{i=c}^{k}\left(\pi_i^{k+1}+\alpha_i\right)=\sum_{i=c}^{k}\pi_i^{k+1}+\sum_{i=c}^{k}\alpha_i\le\sum_{i=c'}^{k}\pi_i^{k+1}+\sum_{i=1}^{k}\alpha_i=\sum_{i=c'}^{k}\pi_i^{k+1}+\pi_{k+1}^{k+1}=\sum_{i=c'}^{k+1}\pi_i^{k+1}=S_e(X^{k+1})$$
$$\Leftrightarrow S_e(X^k)\le S_e(X^{k+1})$$

Given that the quantity π_i^{k+1} is a non-negative number as a probability, the following inequality $\sum_{i=c}^{k}\pi_i^{k+1}\le\sum_{i=c'}^{k}\pi_i^{k+1}$ applies and thus 2.5 is established.

–**Case 3**: $c'>c>1$ The following difference is considered:

$$\begin{aligned}S_e(X^{k+1})-S_e(X^k)&=\sum_{i=c'}^{k+1}\pi_i^{k+1}-\sum_{i=c}^{k}\pi_i^k=\sum_{i=c'}^{k}\pi_i^{k+1}+\pi_{k+1}^{k+1}-\left(\sum_{i=c}^{c'-1}\pi_i^k+\sum_{i=c'}^{k}\pi_i^k\right)\\&=\sum_{i=c'}^{k}\left(\pi_i^{k+1}-\pi_i^k\right)+\pi_{k+1}^{k+1}-\sum_{i=c}^{c'-1}\pi_i^k=\sum_{i=c'}^{k}(-\alpha_i)+\sum_{i=1}^{k}\alpha_i-\sum_{i=c}^{c'-1}\pi_i^k=\sum_{i=1}^{c'-1}\alpha_i-\sum_{i=c}^{c'-1}\pi_i^k\\&=\sum_{i=1}^{c-1}\alpha_i+\sum_{i=c}^{c'-1}\left(\alpha_i-\pi_i^k\right)=\sum_{i=1}^{c-1}\alpha_i+\sum_{i=c}^{c'-1}\left(-\pi_i^{k+1}\right)=\sum_{i=1}^{c-1}\alpha_i-\sum_{i=c}^{c'-1}\pi_i^{k+1}\overset{(c_1)}{\ge}0\end{aligned}$$
$$\Leftrightarrow S_e(X^k)\le S_e(X^{k+1})$$

And thus 2.5 has been established, under condition 1. □

*Remark.*In case 2 the inequality that given a probability space (Ω,F,P), if $A\subseteq B$ then $P(A)\le P(B)$, has been used.
The sensitivity, $S_e(X^k)$, is a non-decreasing function of k, under condition 1.

3 The Case of Uniform Distribution - One Variable

In this section the special case where the discrete variable X^k follows the uniform distribution is presented. Therefore, two discrete random variables X^k and X^{k+1} are considered with k and $k+1$ partitions, respectively. The conditional probabilities π_i^k and π_i^{k+1}, where the discrete random variables X^k and X^{k+1} take a fixed value i, for an individual of the diseased population $(Y=1)$, are given by the following types:

$$\pi_i^k=P(X^k=i\mid Y=1)=\frac{1}{k},\quad i=1,2,\ldots,k,\quad k\in\mathbb{N},k>1$$

and

$$\pi_i^{k+1}=P(X^{k+1}=i\mid Y=1)=\frac{1}{k+1},\quad i=1,2,\ldots,k,k+1,\quad k\in\mathbb{N},k>1.$$

Therefore, according to definitions 2.3 and 2.4 sensitivity, for each case, is equal to:

$$S_e(X^k)=\sum_{i=c}^{k}\pi_i^k=\frac{k-c+1}{k}\tag{3.1}$$

and

$$S_e(X^{k+1})=\sum_{i=c'}^{k+1}\pi_i^{k+1}=\frac{k-c'+2}{k+1}\tag{3.2}$$

where $c\in R_{X^k}=\{1,2,\ldots,k\}$ and $c'\in R_{X^{k+1}}=\{1,2,\ldots,k,k+1\}$.

Proposition 2. *For any $k \in \mathbb{N}$, the sensitivity of index X^k, is a non-decreasing function of k, in the case that variable X^k is distributed uniformly, under a specific condition (condition 2).*

Proof. According to relations 3.1 and 3.2 it is sufficient to be proved that:

$$k < k+1 \Leftrightarrow S_e(X^k) \leq S_e(X^{k+1}) \Leftrightarrow \frac{k-c+1}{k} \leq \frac{k-c'+2}{k+1} \tag{3.3}$$

Let the difference:

$$\begin{aligned} S_e(X^{k+1}) - S_e(X^k) = \sum_{i=c'}^{k+1} \pi_i^{k+1} - \sum_{i=c}^{k} \pi_i^k &= \frac{k-c'+2}{k+1} - \frac{k-c+1}{k} \\ &= \frac{(k-c'+2)\cdot k - (k-c+1)\cdot(k+1)}{k\cdot(k+1)} \\ &= \frac{k^2-kc'+2k-k^2-k+kc+c-k-1}{k\cdot(k+1)} \\ &= \frac{-kc'+kc+c-1}{k\cdot(k+1)} \end{aligned}$$

$$\Leftrightarrow S_e(X^{k+1}) - S_e(X^k) = \frac{-kc'+kc+c-1}{k\cdot(k+1)} \tag{3.4}$$

A sequence of probabilities' differences α_i, $i = 1,2,\ldots,k$ is considered so as to satisfy the following equality:

$$\alpha_i = \pi_i^k - \pi_i^{k+1} = \frac{1}{k} - \frac{1}{k+1} = \frac{1}{k\cdot(k+1)} \tag{3.5}$$

Condition 2: The sequence of probabilities' differences α_i, $i = 1,2,\ldots,k$ satisfies the following inequality

$$\frac{c}{k} - \frac{c'}{k+1} \geq \frac{1}{k\cdot(k+1)} \tag{c_2}$$

in the case that $c < c'$.

The variable X^k is distributed uniformly and according to condition 1 from section 2,

$$\begin{aligned} \sum_{i=1}^{c-1} \alpha_i \geq \sum_{i=c}^{c'-1} \pi_i^{k+1} \Leftrightarrow \sum_{i=1}^{c-1} \frac{1}{k\cdot(k+1)} \geq \sum_{i=c}^{c'-1} \frac{1}{k+1} \Leftrightarrow \frac{c-1}{k\cdot(k+1)} \geq \frac{c'-c}{k+1} \Leftrightarrow \frac{c-1}{k} \geq c'-c \\ \Leftrightarrow c-1 \geq k\cdot c' - k\cdot c \Leftrightarrow c + k\cdot c - k\cdot c' \geq 1 \\ \Leftrightarrow (k+1)\cdot c - k\cdot c' \geq 1 \Leftrightarrow \frac{c}{k} - \frac{c'}{k+1} \geq \frac{1}{k\cdot(k+1)} \end{aligned}$$

It has been shown that condition c_2, is another expression of condition c_1, which is applied, in the case that the variable X^k is distributed uniformly. Therefore, the cases which are distinguished, according to the relative position between c and c', are the following:

– **Case 1**: $c = c' > 1$

According to 3.4 for this case, the difference $S_e(X^{k+1}) - S_e(X^k)$ becomes:

$$S_e(X^{k+1}) - S_e(X^k) = \frac{c-1}{k\cdot(k+1)} > 0$$

because $c > 1$ and $k\cdot(k+1) > 0$. Thus, for $c = c' \geq 1$, 3.3 has been established, i.e. sensitivity is a non-decreasing function of k.

– **Case 2**: $c > c' > 1$

According to 3.4 for this case, the difference $S_e(X^{k+1}) - S_e(X^k)$ becomes:

$$S_e(X^{k+1}) - S_e(X^k) = \frac{-k\cdot c' + k\cdot c + c - 1}{k\cdot(k+1)} = \frac{(-c'+c)\cdot k + c - 1}{k\cdot(k+1)} > 0$$

because $c > 1$, $c - c' > 0$ and $k\cdot(k+1) > 0$. Thus, 3.3 has been established, i.e. sensitivity is a non-decreasing function of k.

–Case 3: $c' > c > 1$
According to 3.4 for this case, the difference $S_e(X^{k+1}) - S_e(X^k)$ becomes:

$$S_e(X^{k+1}) - S_e(X^k) = \frac{-k \cdot c' + k \cdot c + c - 1}{k \cdot (k+1)} = \frac{(k+1) \cdot c + k \cdot c' - 1}{k \cdot (k+1)} > 0$$

Where $\frac{(k+1) \cdot c + k \cdot c' - 1}{k \cdot (k+1)}$ is a non-negative quantity because of condition 2. The numerator of this fraction becomes:

$$(k+1) \cdot c + k \cdot c' - 1 \geq 0 \Leftrightarrow (k+1) \cdot c + k \cdot c' \geq 1 \Leftrightarrow \frac{c}{k} - \frac{c'}{k+1} \geq \frac{1}{k \cdot (k+1)}$$

And thus 3.3 has been established. □

4 Sum of discrete variables - General Case

The indices mentioned in the introduction and used in practice for predicting the value of a binary health outcome Y that indicates the clinical condition of a person, are usually composites, and often, are the sum of other discrete components variables X_j^k, $j = 1, 2, \ldots, m$. For this reason, the sensitivity of a composite health related index $T_m^k = \sum_{j=1}^m X_j^k$ is often interesting to be examined because many decisions, especially in the field of health, depend on many factors.

In this section it is investigated whether the number of the partitions of each discrete variable X_j^k influences the sensitivity S_e, or not, of the index T_m^k, in the case that the distribution of variables X_j^k, $j = 1, 2, \ldots, m$ is unknown. For this purpose, two variables T_m^k and T_m^{k+1} are considered as follows:

$$T_m^k = X_1^k + X_2^k + \ldots + X_m^k = \sum_{j=1}^m X_j^k \text{ and } T_m^{k+1} = X_1^{k+1} + X_2^{k+1} + \ldots + X_m^{k+1} = \sum_{j=1}^m X_j^{k+1}$$

The variables $X_1^k, X_2^k, \ldots, X_m^k$ and $X_1^{k+1}, X_2^{k+1}, \ldots, X_m^{k+1}$ are considered as independent and follow the same distribution. Therefore, variables T_m^k and T_m^{k+1} take values from the following sets:

$$R_{T_m^k} = \{m, m+1, \ldots, mk\} \text{ and } R_{T_m^{k+1}} = \{m, m+1, \ldots, mk, \ldots, m(k+1)\}$$

Then, the sensitivity functions of the variables T_m^k and T_m^{k+1}, in relation to Y, defined as the following conditional probabilities [10]:

$$S_e(T_m^k) = P(T_m^k > l_{0,m}^k \mid Y = 1) \tag{4.1}$$

and

$$S_e(T_m^{k+1}) = P(T_m^{k+1} > l_{0,m}^{k+1} \mid Y = 1) \tag{4.2}$$

where $l_{0,m}^k$ and $l_{0,m}^{k+1}$ are thresholds selected in an appropriate statistical method and $l_{0,m}^k, l_{0,m}^{k+1} \in \mathbb{R}$.

If $\pi_{m,t}^k$ and $\pi_{m,t}^{k+1}$ imply the conditional probabilities that the random variables T_m^k and T_m^{k+1} take a fixed value t, for an individual of the diseased population $(Y = 1)$ then $\pi_{m,t}^k$ and $\pi_{m,t}^{k+1}$ are given from the following formulas:

$$\pi_{m,t}^k = P(T_m^k = t \mid Y = 1), \quad t = m, m+1, \ldots, mk, \text{ and } k \in \mathbb{N}, k > 1.$$

and

$$\pi_{m,t}^{k+1} = P(T_m^{k+1} = t \mid Y = 1), \quad t = m, m+1, \ldots, mk, \ldots, m(k+1), \text{ and } k \in \mathbb{N}, k > 1.$$

Therefore, according to definitions 4.1 and 4.2 sensitivity's functions of variables T_m^k and T_m^{k+1} are given by the formulas 4.3 and 4.4:

$$S_e(T_m^k) = P(T_m^k \geq l_m^k \mid Y = 1) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^k \tag{4.3}$$

$$S_e(T_m^{k+1}) = P(T_m^{k+1} \geq l_m^{k+1} \mid Y = 1) = \sum_{t=l_m^{k+1}}^{m(k+1)} \pi_{m,t}^{k+1} \tag{4.4}$$

where $l_m^k \in R_{T_m^k} = \{m, m+1, \ldots, mk\}$ is the lowest value of T_m^k greater than $l_{0,m}^k$ and $l_m^{k+1} \in R_{T_m^{k+1}} = \{m, m+1, \ldots, mk, \ldots, m(k+1)\}$ is the lowest value of T_m^{k+1} greater than $l_{0,m}^{k+1}$.

Proposition 3. *For any $k \in \mathbb{N}$, the sensitivity of index T_m^k, is a non-decreasing function of k, under a specific condition (condition 3).*

Proof. In order the above proposition to be proved, it is sufficient to be shown that for any $k \in \mathbb{N}$ the following inequality applies:

$$k < k+1 \Leftrightarrow S_e(T_m^k) \leq S_e(T_m^{k+1}) \tag{4.5}$$

According to definitions 4.3 and 4.4, relation 4.5 becomes:

$$\sum_{t=l_m^k}^{mk} \pi_{m,t}^k \leq \sum_{t=l_m^{k+1}}^{m(k+1)} \pi_{m,t}^{k+1} \tag{4.6}$$

A sequence of probabilities' differences $\alpha_{m,t}$, $t = m, m+1, \ldots, mk$ is considered so as to satisfy the following equality:

$$\alpha_{m,t} = \pi_{m,t}^k - \pi_{m,t}^{k+1} \tag{4.7}$$

Condition 3: The sequence of probabilities' differences $\alpha_{m,t}$, $t = m, m+1, \ldots, mk$ satisfies the following inequality

$$\sum_{t=m}^{l_m^k-1} a_{m,t} \geq \sum_{t=l_m^k}^{l_m^{k+1}-1} \pi_{m,t}^{k+1} \geq 0 \tag{c_3}$$

in the case that $l_m^k < l_m^{k+1}$.

The interpretation of condition c_3 is similar to the interpretation of condition c_1, except that the probabilities $\pi_{m,t}^k$ and $\pi_{m,t}^{k+1}$ refer to the variables T_m^k and T_m^{k+1}. More specifically, the sum on the left side of c_3 refers to the difference between the probabilities of the false negative values of indices T_m^k and T_m^{k+1}. On the other hand, the sum on the right side of c_3 refers to the probabilities of the false negative values of index T_m^{k+1}, between thresholds l_m^k and l_m^{k+1}. Practically, condition 3 declares that the probabilities' differences of false negative values, between the index with fewer partitions and the index with more partitions, cumulatively, should be greater than or equal to the probabilities' sum of false negative values of the index T_m^{k+1}, between thresholds l_m^k and l_m^{k+1}.

Therefore, the cases which are distinguished, according to the relative position, between l_m^k and l_m^{k+1} are the following:

–**Case 1**: $l_m^k = l_m^{k+1} > m$

According to relation 4.7, the sensitivity of variable T_m^k becomes:

$$\begin{aligned} S_e(T_m^k) &= \sum_{t=l_m^k}^{mk} \pi_{m,t}^k = \sum_{t=l_m^k}^{mk} \left(\pi_{m,t}^{k+1} + \alpha_{m,t}\right) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} + \sum_{t=l_m^k}^{mk} \alpha_{m,t} \\ &\leq \sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} + \sum_{t=m}^{l_m^k-1} \alpha_{m,t} + \sum_{t=l_m^k}^{mk} \alpha_{m,t} = \sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} + \sum_{t=m}^{mk} \alpha_{m,t} \stackrel{(l_2)}{=} \sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} + \sum_{t=mk+1}^{m(k+1)} \pi_{m,t}^{k+1} = \sum_{t=l_m^k}^{m(k+1)} \pi_{m,t}^{k+1} = S_e(T_m^{k+1}) \\ &\Leftrightarrow S_e(T_m^k) \leq S_e(T_m^{k+1}) \end{aligned}$$

And thus 4.5 is established. □

Lemma 2. *For the sequence of probabilities' differences $\alpha_{m,t}$, $t = m, m+1, \ldots, mk$ applies:*

$$\sum_{t=m}^{mk} \alpha_{m,t} = \sum_{t=mk+1}^{m(k+1)} \pi_{m,t}^{k+1} \tag{l_2}$$

Proof. By summing from $t = m$ to mk both sides in relation 4.7 imply that:

$$\begin{aligned} \alpha_{m,t} = \pi_{m,t}^k - \pi_{m,t}^{k+1} &\Leftrightarrow \sum_{t=m}^{mk} \alpha_{m,t} = \sum_{t=m}^{mk} \left(\pi_{m,t}^k - \pi_{m,t}^{k+1}\right) \Leftrightarrow \sum_{t=m}^{mk} \alpha_{m,t} = \sum_{t=m}^{mk} \pi_{m,t}^k - \sum_{t=m}^{mk} \pi_{m,t}^{k+1} \\ &\Leftrightarrow \sum_{t=m}^{mk} \alpha_{m,t} = 1 - \sum_{t=m}^{mk} \pi_{m,t}^{k+1} \Leftrightarrow \sum_{t=m}^{mk} \alpha_{m,t} = \sum_{t=m}^{m(k+1)} \pi_{m,t}^{k+1} - \sum_{t=m}^{mk} \pi_{m,t}^{k+1} \Leftrightarrow \sum_{t=m}^{mk} \alpha_{m,t} = \sum_{t=mk+1}^{m(k+1)} \pi_{m,t}^{k+1} \end{aligned}$$

–**Case 2**: $l_m^k > l_m^{k+1} > m$
According to relation 4.7, the sensitivity of variable T_m^k becomes:

$$S_e(T_m^k) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^k = \sum_{t=l_m^k}^{mk} \left(\pi_{m,t}^{k+1} + \alpha_{m,t}\right) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} + \sum_{t=l_m^k}^{mk} \alpha_{m,t}$$

$$\leq \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=m}^{l_m^k-1} \alpha_{m,t} + \sum_{t=l_m^k}^{mk} \alpha_{m,t} = \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=m}^{mk} \alpha_{m,t} \overset{(l_2)}{=} \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=mk+1}^{m(k+1)} \pi_{m,t}^{k+1} = \sum_{t=l_m^{k+1}}^{m(k+1)} \pi_{m,t}^{k+1} = S_e(T_m^{k+1})$$

$$\Leftrightarrow S_e(T_m^k) \leq S_e(T_m^{k+1})$$

Given that the quantity $\pi_{m,t}^{k+1}$ is a non-negative number as a probability, the following inequality

$$\sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} \leq \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1}$$

applies and thus 4.5 is established.
–**Case 3**: $l_m^{k+1} > l_m^k > m$
According to relation 4.7, the sensitivity of variable T_m^k becomes:

$$S_e(T_m^k) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^k = \sum_{t=l_m^k}^{mk} \left(\pi_{m,t}^{k+1} + \alpha_{m,t}\right) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^{k+1} + \sum_{t=l_m^k}^{mk} \alpha_{m,t} = \sum_{t=l_m^k}^{l_m^{k+1}-1} \pi_{m,t}^{k+1} + \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=l_m^k}^{mk} \alpha_{m,t}$$

$$\leq \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=m}^{l_m^k-1} \alpha_{m,t} + \sum_{t=l_m^k}^{mk} \alpha_{m,t} = \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=m}^{mk} \alpha_{m,t} \overset{(l_2)}{=} \sum_{t=l_m^{k+1}}^{mk} \pi_{m,t}^{k+1} + \sum_{t=mk+1}^{m(k+1)} \pi_{m,t}^{k+1} = \sum_{t=l_m^{k+1}}^{m(k+1)} \pi_{m,t}^{k+1} = S_e(T_m^{k+1})$$

$$\Leftrightarrow S_e(T_m^k) \leq S_e(T_m^{k+1})$$

And thus 4.5 has been established, under condition 3. □

*Remark.*In case 2 the inequality that given a probability space (Ω, F, P), if $A \subseteq B$ then $P(A) \leq P(B)$, has been used.

5 Sum of discrete variables - The Case of Uniform Distribution

In this section it is shown whether the number of partitions of each discrete variable X_j^k, $j = 1, 2, \ldots, m$ influences the sensitivity of a composite index $T_m^k = \sum_{j=1}^m X_j^k$, in the case that the distribution of variables $X_1^k, X_2^k, \ldots, X_m^k$ is the discrete uniform. In that case the conditional probabilities $\pi_{m,t}^k = P(T_m^k = t \mid Y = 1)$ and $\pi_{m,t}^{k+1} = P(T_m^{k+1} = t | Y = 1)$ of variables T_m^k and T_m^{k+1} become [11,12]:

$$\pi_{m,t}^k = P(T_m^k = t \mid Y = 1) = \frac{1}{k^m} \sum_{r=0}^{\left[\frac{t-m}{k}\right]} (-1)^r \binom{m}{r} \binom{t-kr-1}{m-1}$$

and

$$\pi_{m,t}^{k+1} = P(T_m^{k+1} = t \mid Y = 1) = \frac{1}{(k+1)^m} \sum_{r=0}^{\left[\frac{t-m}{k+1}\right]} (-1)^r \binom{m}{r} \binom{t-(k+1)r-1}{m-1}$$

A sequence of probabilities' differences $\alpha_{m,t}$, $t = m, m+1, \ldots, mk$ is considered so as to satisfy the following equality:

$$\alpha_{m,t} = \pi_{m,t}^k - \pi_{m,t}^{k+1} \tag{5.1}$$

The sequence of probabilities' differences $\alpha_{m,t}$ satisfies the condition c_3, from the previous section, in which the m variables follow an unknown distribution. In the case that the distribution of variables $X_1^k, X_2^k, \ldots, X_m^k$ is the discrete uniform, then the sensitivity functions of the variables T_m^k and T_m^{k+1} become:

$$S_e(T_m^k) = \sum_{t=l_m^k}^{mk} \pi_{m,t}^k = \sum_{t=l_m^k}^{mk} \left\{ \frac{1}{k^m} \sum_{r=0}^{\left[\frac{t-m}{k}\right]} (-1)^r \binom{m}{r} \binom{t-kr-1}{m-1} \right\} \tag{5.2}$$

and

$$S_e(T_m^{k+1}) = \sum_{t=l_m^{k+1}}^{m(k+1)} \pi_{m,t}^{k+1} = \sum_{t=l_m^{k+1}}^{m(k+1)} \left\{ \frac{1}{(k+1)^m} \sum_{r=0}^{\left[\frac{t-m}{k+1}\right]} (-1)^r \binom{m}{r} \binom{t-(k+1)r-1}{m-1} \right\} \tag{5.3}$$

Therefore, in order the sensitivity function's monotony to be investigated, the sign of the following difference must be examined:

$$S_e(T_m^{k+1}) - S_e(T_m^k) = \sum_{t=l_m^{k+1}}^{m(k+1)} \pi_{m,t}^{k+1} - \sum_{t=l_m^k}^{mk} \pi_{m,t}^k =$$

$$\sum_{t=l_m^{k+1}}^{m(k+1)} \left\{ \frac{1}{(k+1)^m} \sum_{r=0}^{\left[\frac{t-m}{k+1}\right]} (-1)^r \binom{m}{r} \binom{t-(k+1)r-1}{m-1} \right\} - \sum_{t=l_m^k}^{mk} \left\{ \frac{1}{k^m} \sum_{r=0}^{\left[\frac{t-m}{k}\right]} (-1)^r \binom{m}{r} \binom{t-kr-1}{m-1} \right\}$$

Tables 1, 2 and 3 present the values of difference $S_e(T_m^{k+1}) - S_e(T_m^k)$ for three cases of the number of variables (m) and partitions (k). More specifically, the cases for $m=2$ and $k=2$, for $m=3$ and $k=2$ as well as for $m=2$ and $k=3$ have been selected. In other cases, for greater values of m and k, the corresponding tables are not practical enough to be presented. Tables 1, 2 and 3 list the values of differences $S_e(T_2^3) - S_e(T_2^2)$, $S_e(T_3^3) - S_e(T_3^2)$ and $S_e(T_2^4) - S_e(T_2^3)$, respectively.

Table 1: Values of difference $S_e(T_2^3) - S_e(T_2^2)$ ($m=2$ and $k=2$) for all the possible values of the thresholds l_2^2 and l_2^3 (For the thresholds applies $l_2^2, l_2^3 > 2$)

$l_2^2 = l_2^3$	$l_2^2 = 3$	$l_2^3 = 3$	$S_e(T_2^3) - S_e(T_2^2) = 0.1389$
	$l_2^2 = 4$	$l_2^3 = 4$	$S_e(T_2^3) - S_e(T_2^2) = 0.4167$
$l_2^2 > l_2^3$	$l_2^2 = 4$	$l_2^3 = 3$	$S_e(T_2^3) - S_e(T_2^2) = 0.6389$
$l_2^2 < l_2^3$	$l_2^2 = 3$	$l_2^3 = 4$	$S_e(T_2^3) - S_e(T_2^2) = -0.0833$
	$l_2^2 = 3$	$l_2^3 = 5$	$S_e(T_2^3) - S_e(T_2^2) = -0.4167$
	$l_2^2 = 3$	$l_2^3 = 6$	$S_e(T_2^3) - S_e(T_2^2) = -0.6389$
	$l_2^2 = 4$	$l_2^3 = 5$	$S_e(T_2^3) - S_e(T_2^2) = 0.0833 > 0$
	$l_2^2 = 4$	$l_2^3 = 6$	$S_e(T_2^3) - S_e(T_2^2) = -0.1389$

Table 2: Values of difference $S_e(T_3^3)-S_e(T_3^2)$ ($m=3$ and $k=2$) for all the possible values of the thresholds l_3^2 and l_3^3 (For the thresholds applies $l_3^2, l_3^3 > 3$)

$l_3^2 = l_3^3$	$l_3^2=4$	$l_3^3=4$	$S_e(T_3^3)-S_e(T_3^2)=0.0880$
	$l_3^2=5$	$l_3^3=5$	$S_e(T_3^3)-S_e(T_3^2)=0.3519$
	$l_3^2=6$	$l_3^3=6$	$S_e(T_3^3)-S_e(T_3^2)=0.5046$
$l_3^2 > l_3^3$	$l_3^2=5$	$l_3^3=4$	$S_e(T_3^3)-S_e(T_3^2)=0.4630$
	$l_3^2=6$	$l_3^3=4$	$S_e(T_3^3)-S_e(T_3^2)=0.8380$
	$l_3^2=6$	$l_3^3=5$	$S_e(T_3^3)-S_e(T_3^2)=0.7269$
$l_3^2 < l_3^3$	$l_3^2=4$	$l_3^3=5$	$S_e(T_3^3)-S_e(T_3^2)=-0.0231$
	$l_3^2=4$	$l_3^3=6$	$S_e(T_3^3)-S_e(T_3^2)=-0.2454$
	$l_3^2=4$	$l_3^3=7$	$S_e(T_3^3)-S_e(T_3^2)=-0.5046$
	$l_3^2=4$	$l_3^3=8$	$S_e(T_3^3)-S_e(T_3^2)=-0.7269$
	$l_3^2=4$	$l_3^3=9$	$S_e(T_3^3)-S_e(T_3^2)=-0.8380$
	$l_3^2=5$	$l_3^3=6$	$S_e(T_3^3)-S_e(T_3^2)=0.1296>0$
	$l_3^2=5$	$l_3^3=7$	$S_e(T_3^3)-S_e(T_3^2)=-0.1296$
	$l_3^2=5$	$l_3^3=8$	$S_e(T_3^3)-S_e(T_3^2)=-0.3519$
	$l_3^2=5$	$l_3^3=9$	$S_e(T_3^3)-S_e(T_3^2)=-0.4630$
	$l_3^2=6$	$l_3^3=7$	$S_e(T_3^3)-S_e(T_3^2)=0.2454>0$
	$l_3^2=6$	$l_3^3=8$	$S_e(T_3^3)-S_e(T_3^2)=0.0231>0$
	$l_3^2=6$	$l_3^3=9$	$S_e(T_3^3)-S_e(T_3^2)=-0.0880$

Table 3: Values of difference $S_e(T_2^4)-S_e(T_2^3)$ ($m=2$ and $k=3$) for all the possible values of the thresholds l_2^3 and l_2^4 (For the thresholds applies $l_2^3, l_2^4 > 2$)

$l_2^3 = l_2^4$	$l_2^3=3$	$l_2^4=3$	$S_e(T_2^4)-S_e(T_2^3)=0.0486$
	$l_2^3=4$	$l_2^4=4$	$S_e(T_2^4)-S_e(T_2^3)=0.1458$
	$l_2^3=5$	$l_2^4=5$	$S_e(T_2^4)-S_e(T_2^3)=0.2917$
	$l_2^3=6$	$l_2^4=6$	$S_e(T_2^4)-S_e(T_2^3)=0.2639$
$l_2^3 > l_2^4$	$l_2^3=4$	$l_2^4=3$	$S_e(T_2^4)-S_e(T_2^3)=0.2708$
	$l_2^3=5$	$l_2^4=3$	$S_e(T_2^4)-S_e(T_2^3)=0.6042$
	$l_2^3=6$	$l_2^4=3$	$S_e(T_2^4)-S_e(T_2^3)=0.8264$
	$l_2^3=5$	$l_2^4=4$	$S_e(T_2^4)-S_e(T_2^3)=0.4792$
	$l_2^3=6$	$l_2^4=4$	$S_e(T_2^4)-S_e(T_2^3)=0.7014$
	$l_2^3=6$	$l_2^4=5$	$S_e(T_2^4)-S_e(T_2^3)=0.5139$
$l_2^3 < l_2^4$	$l_2^3=3$	$l_2^4=4$	$S_e(T_2^4)-S_e(T_2^3)=-0.0764$
	$l_2^3=3$	$l_2^4=5$	$S_e(T_2^4)-S_e(T_2^3)=-0.2639$
	$l_2^3=3$	$l_2^4=6$	$S_e(T_2^4)-S_e(T_2^3)=-0.5139$
	$l_2^3=3$	$l_2^4=7$	$S_e(T_2^4)-S_e(T_2^3)=-0.7014$
	$l_2^3=3$	$l_2^4=8$	$S_e(T_2^4)-S_e(T_2^3)=-0.8264$
	$l_2^3=4$	$l_2^4=5$	$S_e(T_2^4)-S_e(T_2^3)=-0.0417$
	$l_2^3=4$	$l_2^4=6$	$S_e(T_2^4)-S_e(T_2^3)=-0.2917$
	$l_2^3=4$	$l_2^4=7$	$S_e(T_2^4)-S_e(T_2^3)=-0.4792$
	$l_2^3=4$	$l_2^4=8$	$S_e(T_2^4)-S_e(T_2^3)=-0.6042$
	$l_2^3=5$	$l_2^4=6$	$S_e(T_2^4)-S_e(T_2^3)=0.0417>0$
	$l_2^3=5$	$l_2^4=7$	$S_e(T_2^4)-S_e(T_2^3)=-0.1458$
	$l_2^3=5$	$l_2^4=8$	$S_e(T_2^4)-S_e(T_2^3)=-0.2708$
	$l_2^3=6$	$l_2^4=7$	$S_e(T_2^4)-S_e(T_2^3)=0.0764>0$
	$l_2^3=6$	$l_2^4=8$	$S_e(T_2^4)-S_e(T_2^3)=-0.0486$

In Tables 1, 2 and 3, wherever the thresholds l_m^k and l_m^{k+1} satisfy the relations $l_m^k = l_m^{k+1}$ and $l_m^k > l_m^{k+1}$, all values of the difference $S_e(T_m^{k+1}) - S_e(T_m^k)$ are non-negative. If thresholds l_m^k and l_m^{k+1} satisfy the relation $l_m^k < l_m^{k+1}$, the values of the difference $S_e(T_m^{k+1}) - S_e(T_m^k)$ appear both negative and positive, but mostly negative. In the case that the above difference is negative, then thresholds l_m^k and l_m^{k+1} do not satisfy the condition

$$\sum_{t=m}^{l_m^k-1} \alpha_{m,t} \geq \sum_{t=l_m^k}^{l_m^{k+1}-1} \pi_{m,t}^{k+1} \geq 0$$

whereas the positive value appears when thresholds l_m^k and l_m^{k+1} satisfy the above condition. This remark is in agreement with condition 3 of section 3. That means that in those three examples, in which variables $X_1^k, X_2^k, ..., X_m^k$ follow the uniform distribution, the proposition 3 of section 3 is satisfied.

All the values of Tables 1, 2 and 3 have been produced by using a suitable scientific mathematical program (Matlab).

6 Discussion

Health related indices have been extensively used for research in biomedicine, especially in cases where characteristics of individuals such as attitudes and habits are difficult to be measured directly [1]. Although these indices are extremely important in a wide scientific field, there has not been paid enough attention to establish a methodology for their construction and especially for the selection of the optimal number of partitions of their scale.

Extending previous studies based on simulated and empirical data, in this work, it was proved that the sensitivity of a health related index is a non-decreasing function in relation to the number of its partitions under a specific condition (Propositions 1 and 3). Proposition 1 was also verified in the case that variable X^k follows the uniform distribution whereas Proposition 3 was applied in the case that the index T_m^k is the sum of m variables X_j^k, $j = 1, 2, \ldots, m$ that follow the uniform distribution. The aforementioned Propositions give a further reason for concern in the case that the suitable number of partitions of a health related index has to be chosen.This finding of the study is of particular methodological importance in creating more accurate and reliable health related indices, which are designed to predict various health conditions (clinical diagnosis of diseases without symptoms, psychological disorders, nutritional status) [13,14], as well as in being used for various purposes in social science, biosciences etc [1,15,16]. Therefore, this work's result gives a key element in creating an accurate health related index in order to separate the truly diseased people from the untruly ones. The use of such an index can lead to an appropriate treatment and thereby prolong the lifespan and improve the quality of life, or assess different types of diets. Thus, more efficient public health and other social programs may be formulated with a better management of state resources in the field of health.

In a recent publication, which was based only on simulations, it was revealed that the sensitivity of an index is a non-decreasing function of its scale's partitions [7]. Nevertheless, the previous finding was proved in this work. However, simulation methods have certain limitations. The most fundamental of all is that, when simulated data are used in order to investigate a research hypothesis, the findings are based on specific considerations the simulated environment assumes and they are mostly led by the investigator [2]. In this work, it is established that the sensitivity of an index is a non-decreasing function of the number of partitions used by the discrete variable, under a technical condition (i.e. Condition 1) and also the sensitivity of a composite index which is the sum of m variables is a non-decreasing function of the number of partitions used under a technical condition (i.e. Condition 3).

Moreover, in a recent work by Maroulas and Panagiotakos [8], the research hypothesis tested was whether the number of partitions of a discrete variable affects its sensitivity, by the use of one variable. In this work, the hypothesis tested was the same, not only in the case of one discrete variable (Proposition 1), but also in the case of an index considered as the sum of m discrete variables (Proposition 3). In addition, the current work presents the implementation of Proposition 1, in the case of one variable which follows the uniform distribution, as well as three specific examples in the case that the index is the sum of m variables which are distributed uniformly. Therefore, this study is more extended compared to the aforementioned.

Despite the fact that indices with a small number of partitions in their scale are easy for daily use, according to the result of the mathematical proof of this work, as well as the results of the simulation, they do not achieve high diagnostic accuracy. So, it is preferable to use indices with as many partitions as possible, in order to achieve high diagnostic accuracy [7,8]. However, the excessive increase in the number of the partitions of a health related index may cause practical difficulties in using it and may also cause the substantial problem of misclassification. A study using simulated data showed that the presence of misclassification doesn't affect the aforementioned relation between the number of partitions and the sensitivity of the index. Generally, the misclassification issue can be reduced if the researcher utilizes his clinical experience and selects the optimum number of partitions in accordance with the nature of the characteristic the index is designed to evaluate.

Thus, in the future, what should be investigated, theoretically, is whether there is an appropriate number of partitions, beyond which, the increase in diagnostic accuracy (i.e. sensitivity) of the index, is very small or negligible.

Another issue that should be explored theoretically is the correlation between the components which generate the index, as well as the correlation between each component and the health outcome, which the index intends to assess. In addition, a common phenomenon is that some components of the index are correlated with the health outcome more powerfully than some others. Then, it should be further explored whether weights need to be assigned to each component, because all the components of the health related index do not contribute equally to the calculation of the total score [16]. Nevertheless, the presented findings may have a considerable impact on assessing health related behaviors and better exploring the pathophysiological mechanisms of a disease by developing accurate indices that describe human characteristics.

References

[1] Ashima K Kant. Indexes of overall diet quality: a review. *Journal of the American Dietetic Association*, **96**, 785-791, (1996).
[2] Demosthenes Panagiotakos. Health measurement scales: methodological issues. *The open cardiovascular medicine journal*, **3**, 160 (2009).
[3] David L Streiner and Geoffrey R Norman. *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA, (2008).
[4] Douglas N Jackson. A sequential system for personality scale development. *Current topics in clinical and community psychology*, **2**, 61-96 (1970).
[5] Georgia Kourlaba and Demosthenes B Panagiotakos. Dietary quality indices and human health: a review. *Maturitas*, **62**, 18 (2009).
[6] Demosthenes B Panagiotakos, Christos Pitsavos, Fotini Arvaniti, and Christodoulos Stefanadis. Adherence to the mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the meddietscore. *Preventive medicine*, **44**, 335 (2007).
[7] Georgia Kourlaba and Demosthenes Panagiotakos. The diagnostic accuracy of a composite index increases as the number of partitions of the components increases and when specific weights are assigned to each component. *Journal of Applied Statistics*, **37**, 537-554 (2009).
[8] Vasileios Maroulas and Demosthenes B Panagiotakos. Sensitivity of health-related scales is a non-decreasing function of their classes. *arXiv preprint arXiv:1005.3737*, (2010).
[9] Ch A Charalambides. *Enumerative combinatorics*. CRC Press, (2002).
[10] W Daniel Wayne. Biostatistics: a foundation for analysis in the health sciences, (1995).
[11] Charalambos A Charalambides. On the enumeration of certain compositions and related sequences of numbers. *Fibonacci Q*, **20**, 132-146 (1982).
[12] Charalambos A Charalambides. *Combinatorial methods in discrete distributions*, volume 600. Wiley-Interscience, (2005).
[13] Lisa Harnack, Kristin Nicodemus, David R Jacobs, and Aaron R Folsom. An evaluation of the dietary guidelines for americans in relation to cancer occurrence. *The American journal of clinical nutrition*, **76**, 889-896 (2002).
[14] Marjorie L McCullough, Diane Feskanich, Meir J Stampfer, Bernard A Rosner, Frank B Hu, David J Hunter, Jayachandran N Variyam, Graham A Colditz, and Walter C Willett. Adherence to the dietary guidelines for americans and risk of major chronic disease in women. *The American journal of clinical nutrition*, **72**, 1214-1222 (2000).
[15] I Mcdowell. Health measurement scales. *Encyclopedia of Public Health*, (2002).
[16] Sibylle Kranz, Terryl Hartman, AnnaMaria Siega-Riz, Amy H Herring, et al. A diet quality index for american preschoolers based on current dietary intake recommendations and an indicator of energy balance. *Journal of the American Dietetic Association*, **106**, 1594 (2006).