

Original Article

# Estimation of Heterogeneity in Diagnostic Parameters of Age-related Diseases

David Blokh<sup>1</sup> and Ilia Stambler<sup>2\*</sup>

<sup>1</sup>C.D. Technologies Ltd., Israel

<sup>2</sup>Department of Science, Technology and Society, Bar Ilan University, Ramat Gan, Israel

[Received January 4, 2014; Revised February 1, 2014; Accepted February 2, 2014]

**ABSTRACT:** The heterogeneity of parameters is a ubiquitous biological phenomenon, with critical implications for biological systems functioning in normal and diseased states. We developed a method to estimate the level of objects set heterogeneity with reference to particular parameters and applied it to type II diabetes and heart disease, as examples of age-related systemic dysfunctions. The Friedman test was used to establish the existence of heterogeneity. The Newman-Keuls multiple comparison method was used to determine clusters. The normalized Shannon entropy was used to provide the quantitative evaluation of heterogeneity. There was obtained an estimate for the heterogeneity of the diagnostic parameters in healthy subjects, as well as in heart disease and type II diabetes patients, which was strongly related to their age. With aging, as with the diseases, the level of heterogeneity (entropy) was reduced, indicating a formal analogy between these phenomena. The similarity of the patterns in aging and disease suggested a kind of “early aging” of the diseased subjects, or alternatively a “disease-like” aging process, with reference to these particular parameters. The proposed method and its validation on the chronic age-related disease samples may support a way toward a formal mathematical relation between aging and chronic diseases and a formal definition of aging and disease, as determined by particular heterogeneity (entropy) changes.

**Key words:** parameter heterogeneity, Friedman test, Newman-Keuls method, normalized Shannon entropy, diabetes, heart disease, age related disease, aging, system complexity.

An estimate of heterogeneity (contrariness or dissimilarity) of parameters is important for understanding any biological system, and for research of chronic age-related diseases in particular [1-6]. For example, high parameter heterogeneity may completely obscure and confound diagnostic results, while low heterogeneity may reflect a selection bias. Yet, a major difficulty in estimating heterogeneity is that clinical parameters have non-Gaussian distribution. Hence, we propose using the Friedman test to determine the existence or lack of parameters' heterogeneity. In order to separate between clusters of objects described by specific clinical parameters, we use the Newman-Keuls method

for multiple comparisons. This method has been successfully used for the analysis of biomedical and financial data [7-11]. In order to estimate the clustering, i.e. to estimate the degree of partition of the sample into dissimilar groups (the level of heterogeneity), we use the normalized Shannon entropy. This is a dimensionless measure which can permit the concurrent comparison of heterogeneity for a wide variety of different parameters, and in a variety of chronic age-related diseases. Some important applications of the normalized Shannon entropy in medical research have been made earlier [12,13]. Using a similar dimensionless value, i.e. the normalized mutual information, interesting results have been obtained in

\*Correspondence should be addressed to: Ilia Stambler, Department of Science, Technology and Society, Bar Ilan University, Ramat Gan 52900, Israel. Email: [ilia.stambler@gmail.com](mailto:ilia.stambler@gmail.com)

medicine [14], and in particular in oncology [15-17]. The approach proposed in [15], is presented in the monograph [18].

The methods for estimating the heterogeneity, assuming a Gaussian distribution of input parameters, are shown in [19-22]. The present article presents an algorithm for the estimation of heterogeneity, which has no restrictions on the distribution of input parameters and permits to compare the heterogeneity of various objects described by specific parameters. In the present work, this algorithm was applied to investigate the heterogeneity of functional, metabolic and biometric characteristics of healthy subjects and diabetes and heart disease patients as examples of chronic age-related diseases. Thereby an attempt was made to relate parameter heterogeneity with disease status and age. Such an estimation is needed insofar as not just particular, averaged and static parameter values, but the degree of their heterogeneity vary depending on the disease status and the aging process. The estimation of the parameter heterogeneity, essentially reflecting the degree of the system entropy and variability, may thus shed a new light on the underlying processes of aging and disease, providing a unified formal framework for their description. The application of this algorithm may consequently help a diagnosing physician to better evaluate the complexity of the system under examination and hence to better appreciate the complexity of diagnostic results and their relation to age.

## MATERIALS AND METHODS

### Case materials

The proposed method for heterogeneity assessment was validated on datasets for two age-related non-communicable diseases: diabetes and heart disease.

For diabetes, the Pima Indians Diabetes Database of the Johns Hopkins University was analyzed [23]. The representative sample used comprised 161 instances, including healthy subjects and diabetes patients. All the subjects were women at least 21 years old of Pima Indian origin, from Arizona, US. All the cases of diabetes that were included in the database were Type 2 diabetes [24]. The data set included 8 parameters: 1. Number of times pregnant; 2. Plasma glucose concentration at 2 hours in an oral glucose tolerance test; 3. Diastolic blood pressure (mm Hg); 4. Triceps skin fold thickness (mm); 5. 2-hour serum insulin ( $\mu\text{U/ml}$ ); 6. Body mass index (weight in kg/(height in m)<sup>2</sup>); 7. Diabetes pedigree function; 8. Age (years). The current model included the parameters: age, plasma glucose concentration, body mass index, 2 hour serum insulin, and diastolic blood pressure.

For heart disease assessment, we used the Heart Disease Data Set of the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation [25]. The representative sample used comprised 85 instances, including healthy subjects and heart disease patients, aged 34-77. The data set included the following 14 attributes: 1. Age (years); 2. Sex; 3. Chest pain type; 4. Resting blood pressure (in mm Hg on admission to the hospital); 5. Serum cholesterol in mg/dl; 6. Fasting blood sugar (discretized above and below 120 mg/dl); 7. Resting electrocardiographic results (discretized); 8. Maximum heart rate achieved; 9. Exercise induced angina; 10. ST depression induced by exercise relative to rest; 11. The slope of the peak exercise ST segment; 12. Number of major vessels colored by flouroscopy (discretized); 13. Thallium heart scan (normal, fixed defect, reversible defect); 14. The predicted attribute - diagnosis of heart disease (angiographic disease status) with value 0 for diameter narrowing < 50% and value 1 for diameter narrowing > 50%, in any major vessel. The current model included the continuous parameters: age, resting blood pressure, serum cholesterol, and maximum heart rate achieved.

### Algorithm for the assessment of heterogeneity

#### Statement of the problem:

Let us consider that the initial data on  $n$  objects are presented in the form of a  $m \times n$  array  $[a_{ij}]$ , where each column  $j$  is an object described by  $m$  continuous parameters. We need to estimate the heterogeneity of a set, comprised of  $n$  objects, presented by columns of the array  $[a_{ij}]$ .

#### Description of the algorithm:

The algorithm consists of three procedures:

- 1) Determining the heterogeneity using the Friedman test;
- 2) Separating clusters using the Newman-Keuls multiple comparison method;
- 3) Estimating the heterogeneity of the set using the normalized Shannon entropy.

1. For each row of the array  $[a_{ij}]$  we rank its elements and assign rank 1 to the smallest element of the row. We receive the array  $m \times n$  of ranks  $[r_{ij}]$ , where each row of the array contains ranks from 1 до  $n$ . We shall apply the Friedman test to the array  $[r_{ij}]$  [26]. If the Friedman test demonstrates the existence of the objects' heterogeneity, we proceed to the second procedure.

**Table 1.** Sample parameter values for 10 healthy subjects, from the diabetes dataset, aged 21-25

Parameters	Subjects									
	1	2	3	4	5	6	7	8	9	10
Plasma glucose concentration	129	97	88	103	105	141	95	89	142	71
2-Hour serum insulin	270	140	54	82	142	128	38	94	64	76
Diastolic blood pressure	86	66	58	80	64	58	66	66	82	48
Body mass index	35.1	23.2	24.8	19.4	41.5	25.4	19.6	28.1	24.7	20.4

2. We shall further present each object  $j$  by a sum of ranks of its corresponding column  $j$  of the array  $[r_{ij}]$ . We compare the objects, the sums of ranks, by the Newman-Keuls method [7] and obtain the clustering of the set of objects.

3. Let  $X_1, X_2, \dots, X_k$   $1 \leq k \leq n$  be the clustering of the set of objects and  $|X_i|$  the number of elements in the set  $X_i$ .

We shall estimate the heterogeneity of the set of objects (the heterogeneity of parameters describing these objects) using the normalized Shannon entropy [27, 28]:

$$S = -\frac{1}{\ln n} \sum_{i=1}^k (|X_i|/n) \ln(|X_i|/n)$$

Properties of the normalized Shannon entropy:

- 1)  $0 \leq S \leq 1$ ;
- 2)  $S=0$  if and only if the only cluster is the initial set (there is no heterogeneity);
- 3)  $S=1$  if and only if every element of the initial set is a cluster (the maximal heterogeneity). Simply put, zero value of the Shannon entropy would indicate a complete identity of all the elements (complete homogeneity), while the value of 1 would indicate that any individual element has nothing in common with any other (complete heterogeneity).

If the Friedman test does not reject the hypothesis of the absence of row (column) effects, i.e. shows the absence of heterogeneity, then we assume  $S=0$ .

## Results

We demonstrated the applicability of the proposed method for heterogeneity assessment in two age-related diseases: diabetes and heart disease, showing a common pattern of entropy change in aging and different aging-related diseases.

First, to illustrate the proposed approach, we use partial data from the diabetes data set concerning 10 healthy individuals aged 21-25 years. Table 1 shows the initial results for particular parameter values for the 10 subjects, namely: plasma glucose concentration, 2 hour serum insulin, diastolic blood pressure, body mass index. Usually, high values of these parameters are considered as risk factors for diabetes.

Next, we perform consecutively the three procedures of the heterogeneity estimation algorithm:

1. We rank the rows in Table 1 and receive the array  $[r_{ij}]$ . The results are shown in Table 2.

The last row of Table 2 is the sum of ranks of the corresponding columns. Thus, Subject 10 had some of the smallest parameter values (the “best” values or the least associated with diabetes risk) – the plasma glucose concentration was 71 (the smallest value of this parameter in this set, giving it rank 1), 2 hour serum insulin was 76 (rank 4 for insulin), diastolic blood pressure was 48 (rank 1) and the body mass index was 20.4 (rank 3). The overall sum of ranks for this patient was 9 (the smallest sum of ranks for this set, indicating some of the best values in terms of diabetes risk). In contrast, Subject 1 had some of the largest (“worst”) parameter values: plasma glucose concentration was 129 (rank 8 for glucose), 2 hour serum insulin was 270 (the largest value for insulin, giving it rank 10), diastolic blood pressure was 86 (rank 10), and the body mass index was 35.1 (rank 9). The total sum of ranks for this patient was 37 (the highest sum of ranks, indicating some of the worst overall risk factor values).

Let us consider Table 2 as the Friedman statistical model [26], and examine the column effect of this table.

Hypotheses:

$H_0$ : There is no column effect (“null hypothesis”).

$H_1$ : The null hypothesis is invalid.

*Critical range.* The sample is “large”, therefore the critical range is the upper 5%-range of  $\chi^2_9$  distribution.

**Table 2.** Ranked entries of each row of the data array, from the diabetes dataset

Parameters	Subjects									
	1	2	3	4	5	6	7	1	9	10
Plasma glucose concentration	8	5	2	6	7	9	4	3	10	1
2-Hour serum insulin	10	8	2	5	9	7	1	6	3	4
Diastolic blood pressure	10	6	2.5	8	4	2.5	6	6	9	1
Body mass index	9	4	6	1	10	7	2	8	5	3
Sum of ranks	37	23	12.5	20	30	25.5	13	23	27	9

Insofar as the Friedman criterion for “large” groups is approximately equal to the  $\chi^2$  criterion, then for the large groups, for the Friedman test assessment, the  $\chi^2$  is used.

Let us calculate the  $\chi^2$ -criterion [26]. We obtain the value  $\chi^2 = 18.34$ . The critical range is  $\chi^2_0 > 16.92$ . Since  $18.34 > 16.92$ , the null hypothesis with respect to Table 2 is rejected. Thus, according to the Friedman test, the column effect is found. Hence, there is a difference between the columns under consideration. In other words, the Friedman test established the existence of heterogeneity (contrariness or dissimilarity) of objects within this set. Hence we can proceed to the further steps to precisely estimate this heterogeneity.

2. For multiple comparisons, we use the Newman-Keuls test [7].

For  $\alpha_T = 0.05$  ( $\alpha_T$  is the probability at least once to erroneously identify differences) we obtain the critical range for the comparison interval 2 equal 3.92 and  $|R_j - R_{j+1}| > 3.92$  where  $R_j$  and  $R_{j+1}$  are the elements of the column “Sum of ranks” in the  $j$ -th and  $(j+1)$ -th rows of Table 3 respectively (in other words,  $R_j$  and  $R_{j+1}$  are the subjects represented by sums of ranks of the corresponding parameter values).

“By the multiple comparisons, we construct the clustering shown in Table 3. The obtained clustering possesses the following properties: a) for two neighboring clusters of Table 3, the smallest element of one cluster and the greatest element of another cluster located nearby are significantly different ( $\alpha_T = 0.05$ ); b) elements belonging to the same cluster do not differ from each other ( $\alpha_T = 0.05$ ).”

Thus, in the sample of 10 subjects under consideration, 3 clusters are established: 1 subject (Subject 1) is found in the first cluster (with the top sum of ranks for the risk factors), the second cluster (with the intermediate sum of ranks for the risk factors) contains 6

subjects, and the third cluster (with the bottom sum of ranks) contains 3 subjects.

3. According to Table 3, the numbers of elements in each cluster are:

$$|X_1| = 1, |X_2| = 6, |X_3| = 3$$

Now we can calculate the normalized Shannon entropy (S) that provides the exact measure of heterogeneity for this set:

$$S = -((1/10) * \ln(1/10) + (6/10) * \ln(6/10) + (3/10) * \ln(3/10)) / \ln 10 = 0.390$$

**Table 3.** Partition of a set of objects, from the diabetes dataset

Object number	Cluster	Subject	Sum of ranks
1	Cluster 1	1	37
2	Cluster 2	5	30
3		9	27
4		6	25.5
5		8	23
6		2	23
7		4	20
8	Cluster 3	7	13
9		3	12.5
10		10	9

Next we consider the entire set consisting of 53 healthy women, from the diabetes dataset, aged 21-25, where each woman is represented by 4 parameters, as described in the Materials and Methods section. After performing the procedures of the algorithm, we obtain the following clustering (see Table 4): The set, consisting of 53 elements (women) was partitioned into 14 clusters. The

largest number of elements in a single cluster was 20, and in 6 clusters there was only 1 element in each cluster. The normalized Shannon entropy of the clustering equals 0.527, which demonstrates high heterogeneity.

**Table 4.** Partition according to clusters, for 53 healthy women aged 21-25, from the diabetes dataset

Number of elements in a cluster	1	2	4	6	7	20
Number of clusters	6	3	2	1	1	1

Next we analyze the slightly older group of 22 healthy women, aged 26-29. The results are shown in Table 5. Four clusters were found, where the largest number of elements in a single cluster was 9, and 5 clusters contained 1 element each. The normalized Shannon entropy of the clustering equals 0.592, indicating high heterogeneity, about the same or just slightly higher than in the former group.

**Table 5.** Partition according to clusters, for 22 healthy women aged 26-29, from the diabetes dataset

Number of elements in a cluster	1	2	3	9
Number of clusters	5	1	2	1

Then we consider the set consisting of 41 healthy women aged 30-39. After repeating the procedures of the algorithm, we obtain the following clustering (see Table 6): The set, consisting of 41 elements (women) was partitioned into 11 clusters. The largest numbers of elements in a single cluster was 9, while 2 clusters had only 1 element each. The normalized Shannon entropy of the clustering equals 0.583, which demonstrates high heterogeneity and is about the same as in the former group.

**Table 6.** Partition according to clusters, for 41 healthy women aged 30-39, from the diabetes dataset

Number of elements in a cluster	1	2	3	4	6	8	9
Number of clusters	2	3	2	1	1	1	1

Thereafter the following sets were considered: 18 healthy subjects aged 40-49, 24 diabetes patients aged 21-25, 18 patients aged 26-29, 34 patients aged 30-39, and 26

patients aged 40-49 years. For these 5 sets, the Friedman tests showed lack of heterogeneity and  $S=0$ .

We further validated the proposed method on another major age-related disease: the heart disease. We considered four groups of subjects who were characterized by the parameters commonly used to diagnose heart disease: resting blood pressure, serum cholesterol, and maximum heart rate achieved. The four groups were: 17 healthy subjects aged 34-49, 21 healthy subjects aged 50-74; 24 heart patients aged 35-49; and 23 heart patients aged 50-77. After applying the algorithm to the set of healthy subjects aged less than 50 years, we obtained the set partition into 3 clusters, containing 11, 5 and 1 elements (see Table 7). The entropy value was 0.285, showing a considerable degree of heterogeneity in the younger healthy individuals. For the three other sets – healthy individuals over 50, and heart disease patients of all ages, both under and over 50 – the Friedman test showed no heterogeneity, and the entropy value  $S=0$ .

**Table 7.** Partition according to clusters, for 17 healthy subjects aged less than 50 years, from the heart disease dataset

Number of elements in a cluster	1	5	11
Number of clusters	1	1	1

## DISCUSSION

Heterogeneity is an intrinsic phenomenon of any biological system or organism, either unicellular or multicellular. Innumerable causes can contribute to heterogeneity, as biological systems are diverse on every level of organization and in every imaginable parameter. Thus, even within a seemingly homogenous cell population, variability can be great, deriving from different cell stages, different dynamic reactions to exposures, different genetic expressions, different activities of sub-cellular elements, etc [15,29]. Within particular organs and tissues, there are different cell compositions, as well as differences of tissue supply by nutrients in different locations (the tissue center and periphery), and great many other factors. Stemming from the heterogeneity at the lower levels of organization, the heterogeneity of the higher levels, such as individuals in a population, is also high. Thus even within an apparently genetically homogenous population, there could be found great heterogeneity with reference to any physiological and biometric parameter.



The disregard of this heterogeneity, the lack of ability to appreciate its value, can lead to simplistic, biased and misleading results in diagnosis. Hence, the ability to estimate precisely the level of heterogeneity of biological and physiological parameters is of great practical importance for the diagnostic and clinic. Yet, often the discussion of heterogeneity is purely qualitative and impressionistic, without any exact quantitative estimate. Here we provide a new convenient method to estimate precisely the value of heterogeneity, based on personal biological and physiological evaluation, applicable to any set of biological and physiological parameters, at any level of organization. The application of such a method can assist biomedical researchers and diagnosing physicians in assessing the complexity and ambiguity of the model systems they have to confront. Its application in the study of aging and aging-related diseases can be of special significance, as it could provide a formal measure precisely and rigorously relating aging and disease in terms of system heterogeneity.

Here the normalized Shannon entropy was chosen as the measure of heterogeneity. As the results for the diabetes dataset show, there were found considerable differences between diseased and healthy subjects, as well as between older and younger subjects, with references to their entropy (heterogeneity) values. Thus, for the healthy women aged 21-25, the normalized Shannon entropy of the clustering was 0.527, for those aged 26-29 it was 0.592, and for those aged 30-39 it was 0.583, which indicated high levels of heterogeneity. In contrast, for the healthy subjects of older age (aged 40-49), and for all the diseased subjects, from the young age of 21 and onward, there was no heterogeneity shown (zero normalized Shannon entropy). Thus it can be seen that both aging and the diabetic disease, in this sample, are characterized by a lower degree of heterogeneity (entropy) with reference to the measured parameters. In other words, the similarity of the patterns in aging and disease suggests a kind of “early aging” of the diseased subjects, or alternatively a “disease-like” aging process, with reference to these particular parameters. The same pattern of similarity was reconfirmed with reference to the heart disease dataset. In the heart disease dataset, only healthy subjects younger than 50 years showed a considerable degree of heterogeneity (the entropy value  $S=0.285$ ). On the other hand, healthy individuals of higher age (over 50), and diseased subjects of any age, showed lack of heterogeneity (zero entropy value). Thus, also in the case of heart disease, the formal analogy between the aging and disease process was suggested. Of course, a heterogeneity-based biometric that would utilize age as a predictor of disease, and conversely utilize disease as a biomarker of aging, will yet require a thorough

investigation on a much larger sample and range of parameters.

One aspect that may need to be taken into account when developing a heterogeneity-based health biometric is the phenomenon of “heterochrony”, the lack of synchronicity of biological processes. As the results show, the difference between healthy subjects in terms of heterogeneity manifested at different ages: at the age of 40 for the diabetes dataset and 50 for the heart disease dataset. This may be explained by the phenomenon of “heterochrony” or different systems aging at different rates [30,31]. Hence it may be necessary always to delimit the system and set of parameters whose heterogeneity is established. Also, the difference may have arisen due to the specificity of the populations under consideration in the two datasets. Yet, the general trend toward decrease of heterogeneity for aging and disease was shown in both cases of diabetes and heart disease, in both datasets.

The notion of entropy has been used earlier to characterize physiological states generally, and aging and disease in particular. Yet, the physiological and biological interpretations of this term ranged widely. Thus, there has been quite the intuitive “belief that age changes are characterized by increasing entropy, which results in the random loss of molecular fidelity, and accumulates to slowly overwhelm maintenance systems” [32]. Yet, it should be emphasized that for a formal description of an aging and/or diseased system, a formal and calculable entropy value should be used, referring to specific parameters and sub-systems, rather than a general and intuitive concept. There have also been more physicalist interpretations. Thus, in the entropy analysis of the EEG Power Spectrum during sleep in younger and older individuals, the sample entropy of the signal was found to be higher in Stage 2 and REM sleep of the elderly, which was associated with an “alertness promoting” increase of brain wave energy frequencies [33]. In other systems, the interpretations may differ. Thus, in the analysis of body balance, via measurements of the center of pressure, higher entropy was suggested to indicate an increase of automatic (non-volitional) control (greater in young individuals with good balance) as well as an increase in chaotic excursions (greater in older individuals with bad balance). Also in the study of body balance, the interpretations of higher entropy after training included “improved stability”, “increased complexity”, “a more self-organized system”, and more [34]. In view of this diversity, it seems impossible to speak of “changes in entropy” with aging and disease generally, but it always seems necessary to delimit the system under discussion, the formal definition of entropy that is being used and the parameters for which the entropy values are calculated.

Yet despite this variety of interpretations, the dominant (perhaps even counterintuitive) physiological interpretation of entropy change has been that aging and disease are generally associated with a decrease of the system entropy, indicating a loss of the system complexity. Thus, the loss of complexity, shown by lower entropy, was suggested for a wide variety of morbid and age-related conditions, ranging from a more regular (predictable) heart rate in heart diseases, to the periodical (predictable) Parkinsonian tremors [35-37]. The loss of complexity (lower system entropy) has been generally associated with a greater regularity and fixation of the physiological system, in other words with a loss of system variability, and hence the loss of its ability to adapt to the changing environment. This loss of the adaptive capability can be seen as a common feature of both disease and aging. Moreover, the loss of complexity, shown by the lower system entropy, has been suggested as a potentially powerful dynamic biomarker of disease and aging and as a potential metrics to test therapeutic interventions (measuring the therapy's ability to restore the entropy levels) [35]. It is necessary to note, however, that even accepting this interpretation, the loss of system complexity may not necessarily be a detrimental phenomenon, but a protective and adaptive one, diminishing the range of the aged and diseased organism's operation and thus preventing its "overheating" (as apparently happens in repair mechanisms of sleep which involve a high degree of synchronicity [38]).

In the present work, the found patterns of heterogeneity in the diabetes and heart disease patients generally conformed with the prevalent findings of entropy loss in aging and diseases. Within the present populations, and with reference to the present parameters, the subjects may have become "more similar" to each other (less heterogeneous) as they aged and succumbed to disease, simply due to the loss of their adaptive function – generally diminishing the range of their reactions to confine it to a narrower framework. It is possible that the heterogeneity diminished due to an execution of some genetic program of development, or due to a protective mechanism to withstand adverse environmental circumstances ("finding safety in a confined space"). Whatever the explanation, in practical terms, the levels of heterogeneity may serve as an additional tool in the epidemiological and nosological arsenal, helping to analyze the incidence of diabetes, heart disease and other age-related diseases and conditions. Following further development, the use of information-theoretical measures, such as entropy, may become a generally applicable formal quantitative tool to describe senescent and diseased states as well as restoration of physiological function

## Conflict of interest

We have no conflict of interest.

## References

- [1] Dussoix P, Vaxillaire M, Iynedjian PB, Tiercy JM, Ruiz J, Spinass GA, et al. (1997). Diagnostic heterogeneity of diabetes in lean young adults. Classification based on immunological and genetic parameters. *Diabetes*, 46:622-631.
- [2] Radtke MA, Midthjell K, Nilsen TI, Grill V (2009). Heterogeneity of patients with latent autoimmune diabetes in adults: Linkage to autoimmunity is apparent only in those with perceived need for insulin treatment. *Diabetes Care*, 32:245-250.
- [3] Bhopal R, Unwin N, White M, Yallop J, Walker L, Alberti KG, et al. (1999). Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. *Brit Med J*, 319:215-220.
- [4] Mestroni L, Rocco C, Gregori D, Sinagra G, Di Lenarda A, Miocic S, et al. (1999). Familial dilated cardiomyopathy: evidence for genetic and phenotypic heterogeneity. *J Am Coll Cardiol*, 34:181-190.
- [5] Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, et al. (2014). Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*, 25:91-101.
- [6] Marjanovic ND, Weinberg RA, Chaffer CL (2013). Cell plasticity and heterogeneity in cancer. *Clin Chem*, 59:168-179.
- [7] Glantz SA. *Primer of Biostatistics*, 4th edition. New York: McGraw-Hill; 1994.
- [8] Wu SS, Wang W, Annis DH (2008). On identification of the number of best treatments using the Newman-Keuls Test. *Biom J*, 50:861-869.
- [9] Blokh D (2013). Information-theory analysis of cell characteristics in breast cancer patients. *Int J Bioinform Biosci*, 3:1-5.
- [10] Blokh D (2012). Clustering financial time series via information-theory analysis and rank statistics. *J Patt Reco Res*, 7:106-115.
- [11] Blokh D (2013). Financial time series analysis based on normalized mutual information functions. *Int J Cybern Inform*, 2:1-8.
- [12] Berretta R, Moscato P (2010). Cancer biomarker discovery: The entropic hallmark. *PLoS One*, 5:e12262.
- [13] Greenier JL, Van Rompay KK, Montefiori D, Earl P, Moss B, Marthas ML (2005). Simian immunodeficiency virus (SIV) envelope quasispecies transmission and evolution in infant rhesus macaques after oral challenge with uncloned SIVmac251: increased diversity is associated with neutralizing antibodies and improved survival in previously immunized animals. *Virol J*, 2:e11.

- [14] Zvarova J, Studeny M (1997). Information theoretical approach to constitution and reduction of medical data. *Int J Med Inform*, 45:65-74.
- [15] Blokh D, Stambler I, Afrimzon E, Shafran Y, Korech E, Sandbank J, et al. (2007). The information-theory analysis of Michaelis–Menten constants for detection of breast cancer. *Cancer Detect Prev*, 31:489-498.
- [16] Blokh D, Zurgil N, Stambler I, Afrimzon E, Shafran Y, Korech E, Sandbank J, Deutsch M (2008). An information-theoretical model for breast cancer detection. *Methods Inf Med*, 47:322-327.
- [17] Blokh D, Stambler I, Afrimzon E, Platkov M, Shafran Y, Korech E, et al. (2009). Comparative analysis of cell parameter groups for breast cancer detection. *Comput Methods Programs Biomed*, 94:239-249.
- [18] Gutierrez Diez PJ, Russo IH, Russo J. *The Evolution of the Use of Mathematics in Cancer Research*. New York: Springer; 2012.
- [19] Lawson AB (1994). Using spatial Gaussian priors to model heterogeneity in environmental epidemiology. *Statistician*, 43:69-76.
- [20] Foulley JL, Quaas RL (1995). Heterogeneous variances in Gaussian linear mixed models. *Genet Sel Evol*, 27:211-228.
- [21] Böhning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A (2002). Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics*, 3:445-457.
- [22] Kanematsu N, Komori M, Yonai S, Ishizaki A (2009). Dynamic splitting of Gaussian pencil beams in heterogeneity-correction algorithms for radiotherapy with heavy charged particles. *Phys Med Biol*, 54:2015-2027.
- [23] UCI Machine Learning Repository. Pima Indians Diabetes Data Set. Original owner: US National Institute of Diabetes and Digestive and Kidney Diseases. Donor of Database: The Johns Hopkins University, Accessed January 2014, Retrieved from: <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>.
- [24] Baier LJ, Hanson RL (2004). Genetic studies of the etiology of type 2 diabetes in Pima Indians. Hunting for pieces to a complicated puzzle. *Diabetes*, 53:1181-1186.
- [25] UCI Machine Learning Repository. Heart Disease Data Set. Creator: V.A. Medical Center, Long Beach and Cleveland Clinic Foundation, Accessed January 2014, Retrieved from: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>.
- [26] Conover WJ. *Practical Nonparametric Statistics*. New York: Wiley-Interscience; 1999.
- [27] Alter O, Brown PO, Botstein D (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA*, 97:10101-10106.
- [28] Rosso OA, Masoller C (2009). Detecting and quantifying temporal correlations in stochastic resonance via information theory measures. *Eur Phys J B*, 69:37-43.
- [29] Elsasser WM (1984). Outline of a theory of cellular heterogeneity. *Proc Natl Acad Sci USA*, 81:5126-5129.
- [30] Lapidot MB (1987). Does the brain age uniformly? Evidence from effects of smooth pursuit eye movements on verbal and visual tasks. *J Gerontol*, 42:329-331.
- [31] Chen D, Bruno J, Easlon E, Lin SJ, Cheng HL, Alt FW, Guarente L (2008). Tissue-specific regulation of SIRT1 by calorie restriction. *Genes Dev*, 22:1753-1757.
- [32] Hayflick L (2007). Entropy explains aging, genetic determinism explains longevity, and undefined terminology explains misunderstanding both. *PLoS Genet*, 3:2351-2354.
- [33] Bruce EN, Bruce MC, Vennelaganti S (2009). Sample entropy tracks changes in EEG power spectrum with sleep state and aging. *J Clin Neurophysiol*, 26:257-266.
- [34] Borg FG, Laxåback G (2010). Entropy of balance – some recent results. *J Neuroeng Rehabil*, 7:e38.
- [35] Lipsitz LA, Goldberger AL (1992). Loss of 'complexity' and aging: potential applications of fractals and chaos theory to senescence. *JAMA*, 267:1806-1809.
- [36] Goldberger AL (1996). Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *Lancet*, 347:1312-1314.
- [37] Goldberger AL, Peng CK, Lipsitz LA (2002). What is physiologic complexity and how does it change with aging and disease? *Neurobiol Aging*, 23:23-26.
- [38] Vyazovskiy LV, Harris KD. Sleep and the single neuron: the role of global slow oscillations in individual cell rest. *Nat Rev Neurosci*, 14:443-451.