STUDIES OF DIVERSITY INDICES ON MORTALITY STATISTICS

by

J. IZSÁK* – P. JUHÁSZ-NAGY**

Summary

The pattern of the combination of diseases or causes of death in a population can be studied by methods used in ecology, e.g. by the study of diversity.

We used as study material the mortality statistics in the 1974 and 1975 Annual of Vital Statistics of the United States. After the calculation of several types of diversity indices age-diversity curves were plotted. On this basis the changes of diversity as a function of age and the sexual differences of diversity are discussed by sections, and the statistical characteristics of the diversity indices are studied in the given study material. Diversity usually increases until the age of $20-30$, then it decreases. H i l l's diversity indices and H u r l b e r t's indices run similarly as a function of age, the shape of the curves of the Q index is different and inconstant. On the basis of our results the rise and fall of diversity by age and the greater diversity in the case of females can be considered as a phenomenon that is independent of the choice of diversity index.

Key Words: mortality structure, diversity index, aging, sexual differences

Introduction

Diversity indices are widespread tools of investigation in ecology, sociobiology, population dynamics and several other biological disciplines. Although the large *epidemiological* material published in taxonomic system offers a good opportunity for utilizing the apparatus of diversity studies, in our knowledge, barely has this been attempted previously. We know only of the fitting the logseries distribution to mortality data (H e r d a n [1]), later, W. T S t i l l e mentions the examination of Shannon's H' statistic (see later) on a morbidity material in the discussion of a presen-

* Teacher's College Szombathely H – 9701, Dep. of Biology
** Eötvös University, Department of Plant Taxonomy and Ecology, Budapest

tation [2]. As the increase of the methods of investigation of epidemiology lags behind the fast increase of epidemiological statistics — causing the relative under-utilization of the mass of data accumulating throughout the world [3,4] — it seems desirable to develop new examination possibilities of epidemiology. We attempted to apply diversity studies in epidemiology for this reason [5, 6]. In the course of these studies, for example, the ,,units" may be formed on the basis of certain diagnoses of diseases of the International Code of Diseases (ICD), while the categories may be formed by the three-digit ICD categories. Diversity studies may play a role in the field of epidemiology to reinforce the global investigation point of view. The analysis of the occurance of certain diseases or restricted ensemble of diseases is not complemented sufficiently today by a global type of investigation to which the cenological approach complementing the description of the presence and range of certain species may serve as a good example. An epidemiological investigation aspect is conceivable according to wich the sum of disease frequencies constitute the universe, i.e. by pushing into the background the epidemiological analysis of the individual diseases, every disease frequency is considered only as one of the many (I z s á k — J u h á s z - N a g y [5]).

The diversity studies performed on the epidemiological material may be advantageous methodologically, too, because a large part of the epidemiological statistics may serve as an easily available reference base for the study of diversity indices.

In our previous studies both on morbidity and mortality materials it was found that the diversity of male groups is usually smaller than that of the corresponding female groups [5, 6]. We also described the characteristic development of mortality diversity according to age on the basis of the US mortality statistics. Diversity was always measured by the relative Brillouin index. However, on the one hand, this index measures only the evenness component of diversity (P e e t [7], P i e l o u [8]), on the other hand it is very sensitive to the dominant frequencies. For this reason we felt it necessary to repeat our studies in part by diversity indices based on different conceptions and having different statistical characteristics.

Study material

Our study was based on the 1974 and 1975 mortality statistics of the USA (VitalStatisticsofthe US 1974 and 1975, [9a, 9b]). The diversity indices were computed for the mortality statistics of whites, negroes and ,,other races", broken down according to sex and age, for disease sections I, II, III, VI, VII, VIII, IX, X and XIV and for the total of the considered categories. The number of mortality categories considered within the individual sections of diseases can be seen in Table 1. The rest of the sections were excluded from the study either because of the small number of categories (sections IV, V, XII, XIII, XV and XVI) or because of the special nature of the section (sections XI and EXVII). Within the studied sections we had to omit categories that appear in the statistical publication in

combination with quite different categories (e.g. category ,,other bacterial diseases" contains the three-digit categories ICD 007, 021, 024 – 027, 031 and 039, etc.). The combined categories, however, were not excluded if the combination concerned only a few and similar three-digit categories (e.g. the combined category ,,tuberculosis of respiratory systems" was not omitted, because it contains categories 010, 011 and 012, that are similar). The four-digit categories were combined within the appropriate three-digit categories. Naturally, the decision concerning the retention or omission of categories contains subjective elements, too, but keeping all of the categories would not have been correct at all.

Table 1

Number of considered categories in the sections (s) and the parameter m of the diversity index s(m)

Disease section	s	m
I. Infective and parasitic diseases	45	30
II. Neoplasms	37 (males)	30
III. Endocrine, nutritional and metabolic diseases	7	–
VI. Diseases of the nervous system and sense organs ...	6	–
VII. Diseases of the circulatory system	29	20
VIII. Diseases of the respiratory system	16	5
IX. Diseases of the digestive system	11	4
X. Diseases of the genitourinary system	5	–
XIV. Congenital anomalies	5	–
Total	161 (males) 165 (females)	10 and 100 10 and 100

The changes of the different diversity indices according to age are shown on diversity graphs (Figures 1 – 4). The horizontal axis of the graphs shows the five-year age groups. (The first age group (1 – 4 years) includes only four years because infant mortality has been excluded from the group of 0 – 4 years.)

Diversity indices used in this paper

Subsequently, n will be the number of elements in the studied (sub-) population, s will be the number of the non-empty categories (in our case that of the three-digit categories of ICD), n_i (i = 1, 2, . . ., s) is the frequency of the i-th category, p_i is the probability of falling in the i-th category in the appropriate multinomial model.

The relative B r i l l o u i n index mentioned in the Introduction is the quotient of the H (absolute) B r i l l o u i n index and the greatest possible H_{max} B r i l l o u i n index belonging to the given n and s: $J = = H/H_{max}$ (P i e l o u [8]. The relative B r i l l o u i n index measures
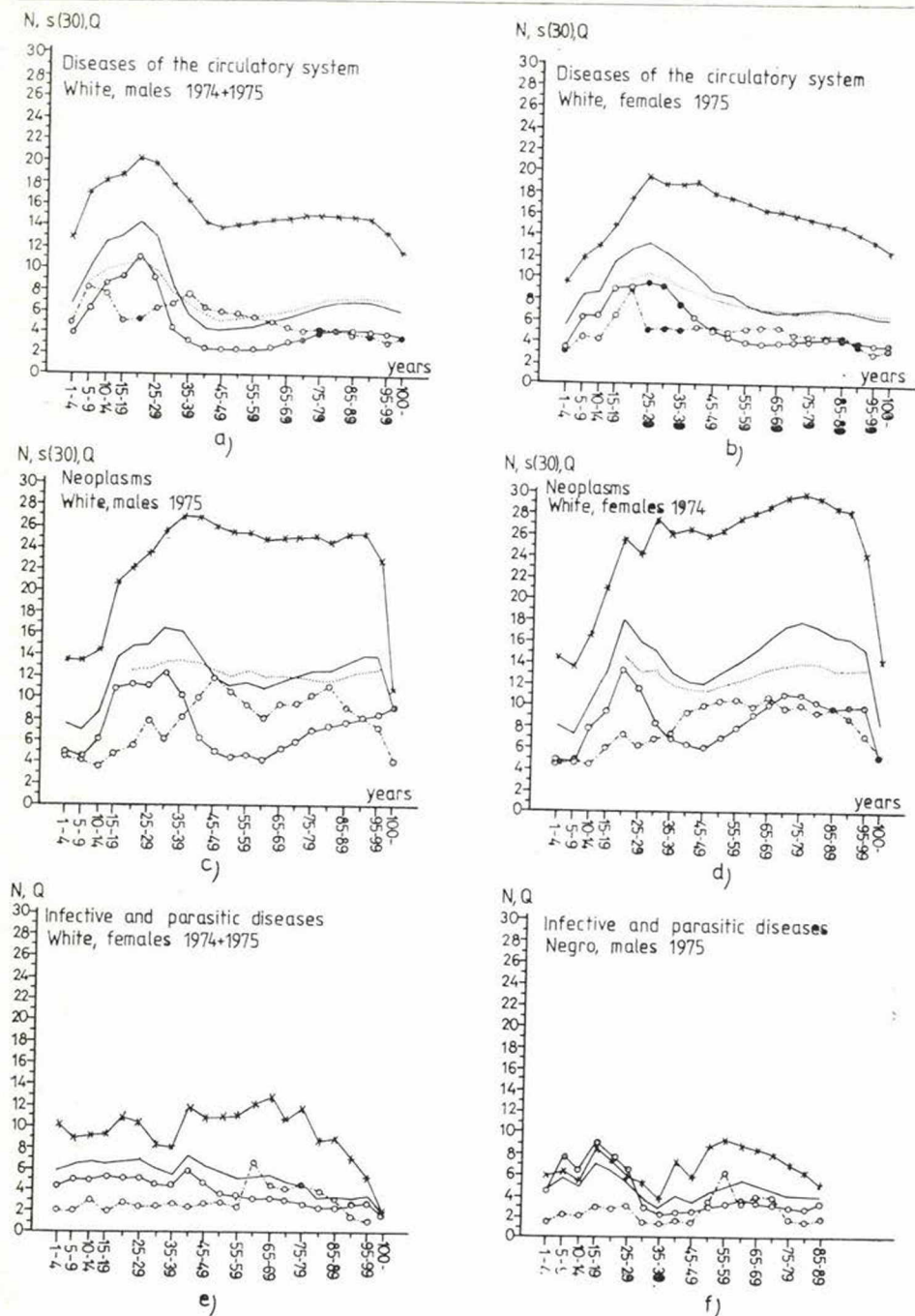
Fig. 1. Diversity curves. Sections I, II, VII. N_2: ○——○, N_1: ——, $N_{1/3}$: ×——×, $s(m)$: ......, Q: ○—·—○
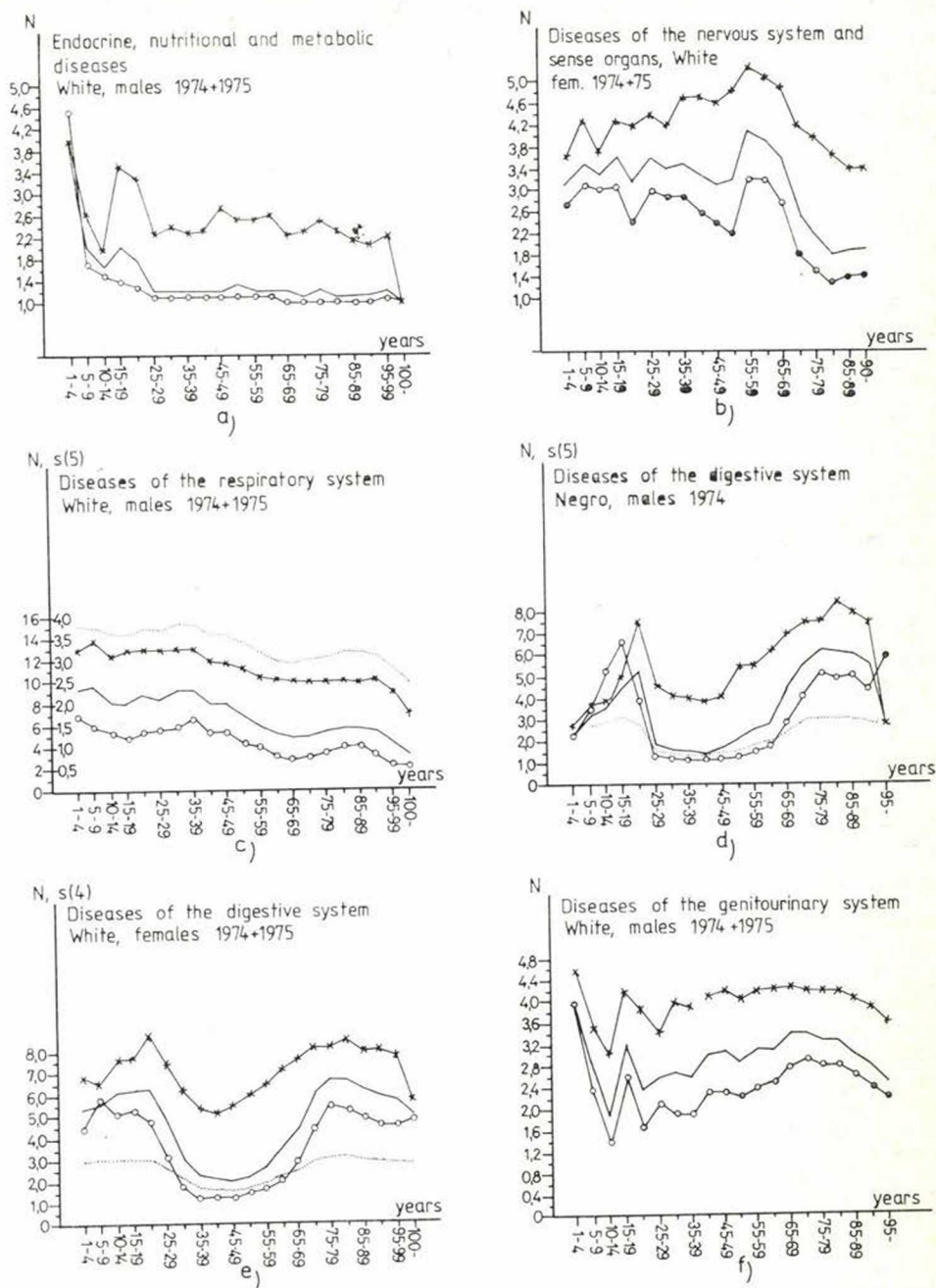
Fig. 2. Diversity curves. Sections III, VI, VIII, IX, X. N_2: o—o , N_1:—— , $N_{1/2}$: ×——× , $s(m)$:······
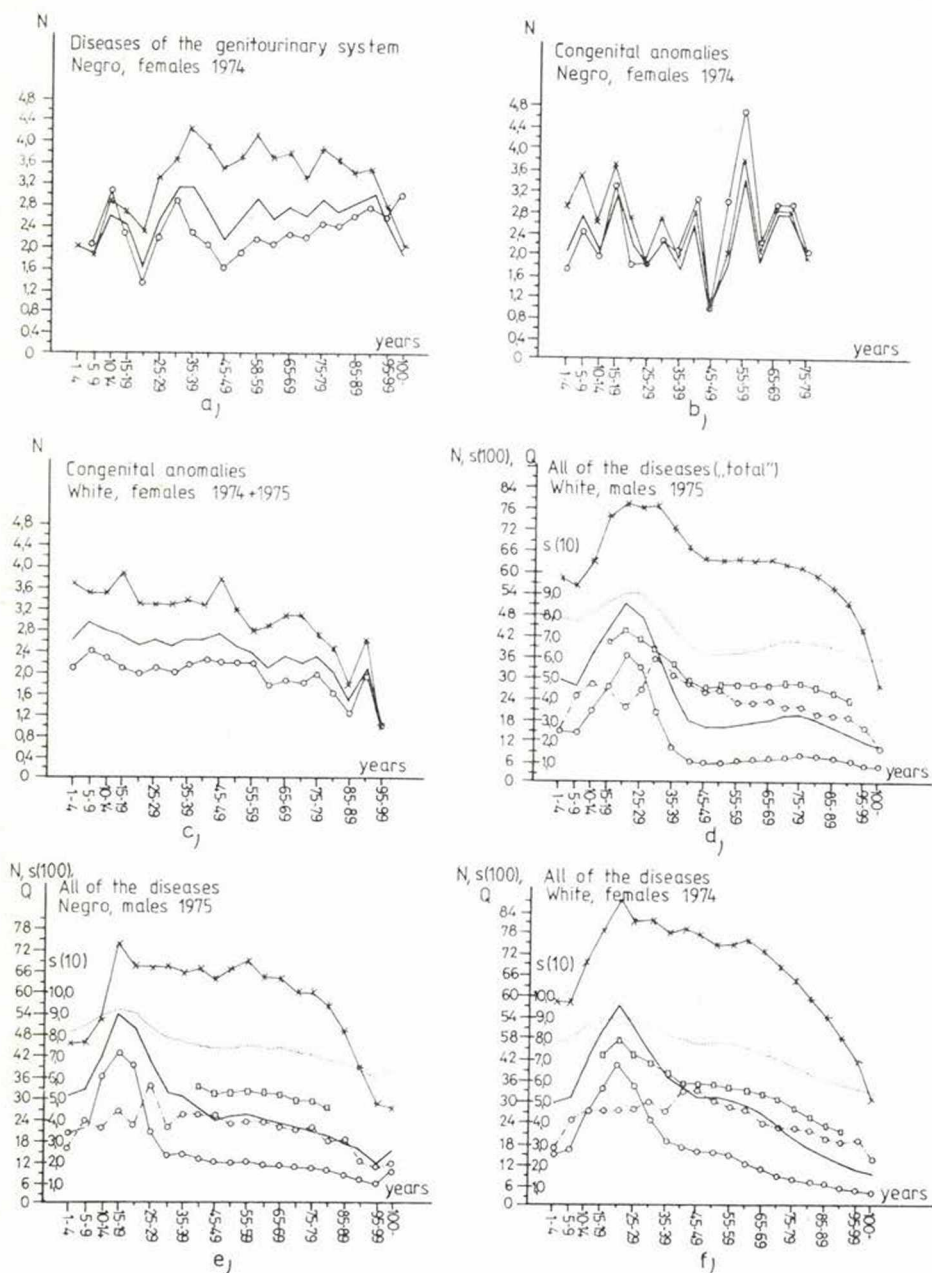
Fig. 3. Diversity curves. Sections X, XIV and combined material ("total groups").
N_2: o——o, N_1:——, $N_{1/3}$: ×——×, s(10): ······· , s(100): □——□, Q:o——·——o
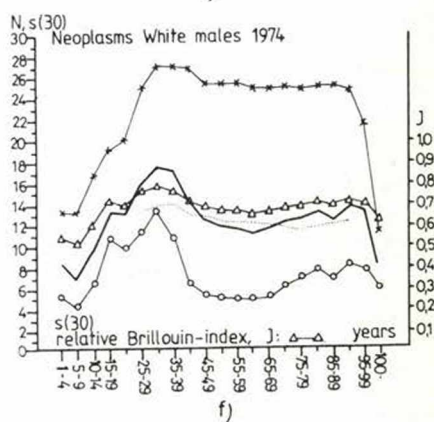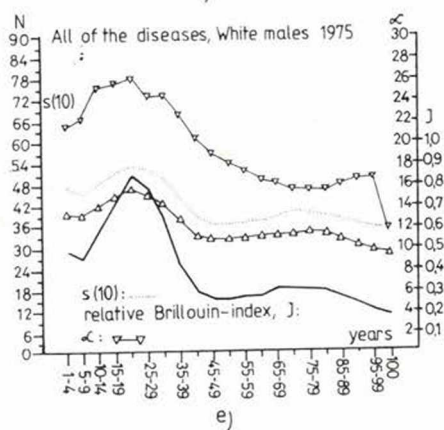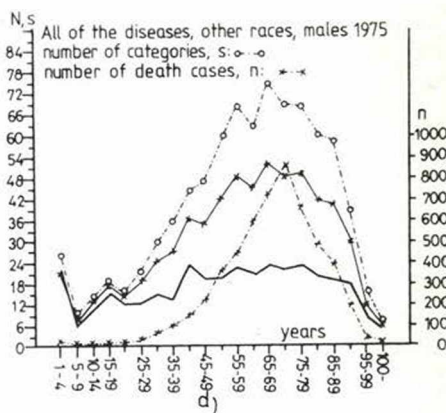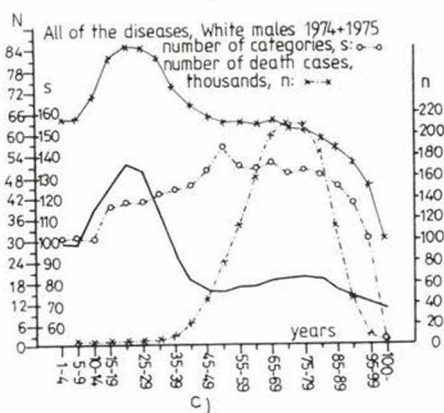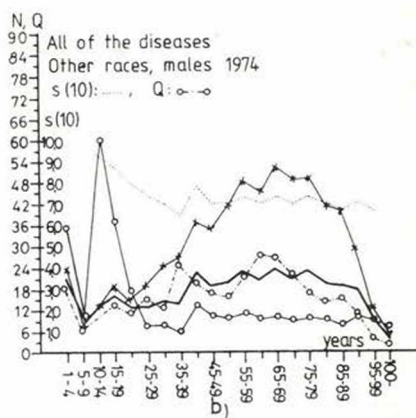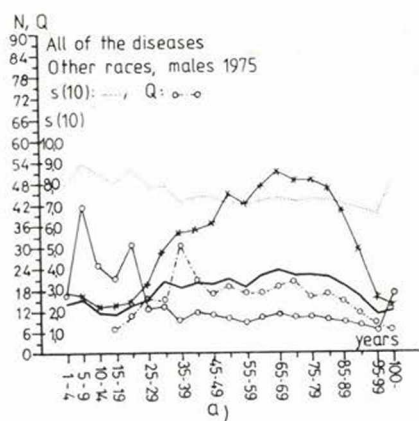
Fig. 4. Diversity curves for the total groups. Above mentioned indices and other parameters. In all cases N_2: o——o , N_1: ——, $N_{1/3}$: ×——×. Other symbols see at the curves.

the component of diversity that arises from the difference between the concrete frequency distribution and the corresponding uniform distribution. For this reason J is listed among the so called evenness indices [7, 8]. The diversity indices to be discussed below, however, also consider the component of diversity that concerns the number of categories.

It is known that the three evident requirements that can be raised for the diversity indices are met essentially only by the B r i l l o u i n index and the related H' S h a n n o n's information index [8]. Due to their other different favourable characteristics, however, several other diversity indices have also become widespread (W i l l i a m s [10], S i m p s o n [11], F a g e r [12], H u r l b e r t [13], H i l l [14], P e e t [7], K e m p-t o n – T a y l o r [15], P a t i l – T a i l l i e [16], etc.).

First the family with one parameter of H i l l's indices are mentioned among the diversity indices used here, that is defined by the following formula (H i l l [14]):

$$N_a = \left(\sum_{i=1}^{s} p_i^a \right)^{1/(1-a)} \qquad a \neq 1, \ N_1 = \lim_{a \to 1} N_a).$$

It is known that as a decreases N_a emphasizes small frequencies more and more, N_a is a strictly decreasing function of a, the maximum of N_a in every case of fixed a is s, and $N_o \equiv s$.

H i l l's family of indices alone is already capable of moderating the role of dominant frequencies. The appropriate estimates of N_2, $N_{1/3}$ and N_1 were calculated from our data.

In the development of N_2 the role of high frequencies is dominant [7, 17]. Its significance is increased by the fact that algebraically it is the same as the invert of the Y S i m p s o n index that is a long used diversity measure [10, 13, 18, 19]. Beside the H i l l-type of interpretation N_2 can also be conceived as a certain equivalence number of species [7]. Its estimate used in this paper is (also see [10, 19]):

$$\hat{N}_2 = \frac{n(n-1)}{\sum_{i=1}^{s} n_i(n_i-1)} .$$

In cases of small number of elements $\hat{N}_2$ proved to be significantly greater than realistic, although $\hat{N}_1^{-1}$ is an unbiased estimate of Y.

$N_{1/3}$ was studied, because this index of diversity strongly emphasizes low frequencies [14]. Its applied estimate is:

$$\hat{N}_{1/3} = \left(\sum_{i=1}^{s} \left(\frac{n_i}{n} \right)^{1/3} \right)^{1/(1-1/3)} .$$

wich underestimates $N_{1/3}$ [17].

It seemed to be useful to calculate N_1, too. On the basis of the mentioned characteristics this index, wich is characterized by emphasizing higher

frequencies (17), is in an intermediate position in this respect between N_2 and $N_{1/3}$. On the basis of the relationship $N_2 = \exp(H')$, N_2 is in a close connection with S h a n n o n's index, in fact, on the basis of the $\lim_{\substack{\min n_i \to \infty}} H =$
$= H'$ relationship it is in connection with the B r i l l o u i n index, too. N_1 can also be interpreted as a certain equivalence number of species (M a c A r t h u r [20], P e e t [7]). Its significance is increased by the fact that due to statistical reasons several authors switch to $\exp(H')$, i.e. to N_1, at the calculation of H' S h a n n o n-index, too [7, 13, 21]. Its applied estimate is:

$$\hat{N}_1 = \exp\left(-\sum_{i=1}^{s} \frac{n_i}{n} \log \frac{n_i}{n}\right),$$

wich underestimates N_1 [22, 2].

We will not deal with the details of the statistical characteristics of the $\hat{N}_a$ estimates, but one of the important, elementary sources of bias is mentioned here: the smallest positive value of n_i can be 1, thus, with the decrease of sample size many singletons disappear and the estimated values of diversity decrease. This effect increases with the decrease of parameter a. E.g. $\hat{N}_{1/3}$ is already very dependent on sample size (see below).

Another conception lies behind the index family

$$s(m) = \sum_{i=1}^{s} \left(1 - (1 - p_i)^m\right) \quad (m = 1, 2, \ldots).$$

Choosing m number of units from the multinomial population (p_1, p_2, ... p_s), the expectation of the number of categories represented by them is $s(m)$ (H u r l b e r t [13], S i m b e r l o f f [23]. With the increase of m, $s(m)$ emphasizes lower frequencies. E.g. in the case of $m = 2$, $s(m)$ is in a direct connection with the S i m p s o n index: $s(2) = 2 - Y$, on the other hand, $\lim_{m \to \infty} s(m) = s$.

Here we can also say that the index family is theoretically suitable for studying and influencing the role of leading frequencies (S m i t h − G r a s s l e [24]). $s(m)$ diversity indices were calculated for sections I, II, VII, VIII and IX, beside parameter values changing by sections (see Table 1). Two parameter values were chosen only in the combined material, namely 10 and 100.

The sample statistic of

$$\hat{s}(m) = \sum_{i=1}^{s} \left(1 - \binom{n - n_i}{m} \middle/ \binom{n}{m}\right),$$

was used for the estimation of $s(m)$, because this is the minimum variance unbiased (MVU) estimate of $s(m)$ [24], wich is evidently a very favourable condition. The statistical characteristics of $s(m)$ are dealt with by e.g. H e c k et al [25], S i m b e r l o f f [26], and S m i t h et al. [27].

Finally, we also used the following index (K e m p t o n – T a y l o r [15]) for the study of diversity:

$$Q = (s/2)/\log(P_{s/4}/P_{3s/4}),$$

were $p_{s/4}$ and $p_{3s/4}$ are the upper and lower quartiles, respectively. (We can only refer to the different models where the p_i-s are the realizations of a random variable [8, 28]. We can talk about quartiles here in this sense.) Thus, this index of diversity disregards completely leading causes of death. We deemed its examination advisable, because its development may stand for the structural and divided change of the combination of frequencies. The examination of Q is justified only in case of a great number of categories. Its applied estimation is:

$$\hat{Q} = (s/2)/\log(n_{s1}/n_{s2}),$$

where n_{s1} is the $\left[\dfrac{s}{4}\right]$ th. while n_{s2} is the $\left[\dfrac{3s}{4}\right]$ th of the n_i frequencies in a decreasing order of size. The bias of $\hat{Q}$ depends primarily on the ratio of all of the categories represented in the sample [19].

Finally, it has to be noted that in some extreme cases, where the number of elements is very small, the above formulas of N_2 and Q are meaningless; this, however, does not cause any confusion in practice. The diversity indices were calculated by a FORTRAN program.

Some authors list indices $s(m)$ and Q among the so called richness indices, while they list the H i l l indices among the so called heterogeneity indices (P e e t [7]). Similarly to several other authors we do not find it possible to differentiate these two groups, although opinions vary in this respect (P e e t [7], K e m p t o n [17], E n g e n [29]). An abundant bibliography can be found in P e e t's earlier [7] and G r a s s l e et al.'s more recent [30] papers on the literature of diversity.

Basic statistical characteristics of the diversity curves

The discussion of the results is mostly based on the phenomenological description of the diversity curves. For easier orientation the statistical and biological phenomena are discussed separately where possible.

Naturally, the statistical characteristics of the diversity indices stand behind that of the curves. We have mentioned already the ones that are the most important from our point of view. There are, however, special problems concerning the study material that are not discussed in the literature on diversity. First of all, the ICD categories are not as well defined categories as the taxonomic categories used in ecology. It cannot be let out of account that there are — probably a lot of — mistakes made at the classification to the causes of death, too. The question of combined categories is also problematic at the use of vital statistics. Relying on all this and considering the originality of the study material in the circle of diversity stu-

dies, as a first step, the review of the elementary statistical observations are put to the fore.

It can be observed in almost all of the figures that the curve of $\hat{N}_{1/3}$ can be found above the curve of $\hat{N}_1$, and that curve $\hat{N}_2$ is below curve $\hat{N}_1$. On the basis of the mentioned characteristics of N_a indices this is natural in the case of $\hat{N}_{1/3}$ and $\hat{N}_1$, because the n_i/n relative frequencies can be considered as certain p'_i ($i = 1, 2, \ldots, s$) probabilities. Because of the overestimation of N_2 by $\hat{N}_2$ — wich is considerable only if the sample size is small — the points of the curve of $\hat{N}_2$ sometimes fall above the points of $\hat{N}_1$, even above that of $\hat{N}_{1/3}$ (e.g. Figs. 1f, 3b and 4a). Another important feature is based on the mentioned characteristics: as N_2 is the most sensitive to outstandingly great frequencies and N_1 is also quite sensitive to them, it can usually be observed that in the case of a given group the fall after a maximum point is the fastest in the case of N_2 and N_1, while in $N_{1/3}$ the hollow after the maximum point may even disappear together with the minimum point (Figs. 1b, 1c, 3d).

The smoothness of the curves may depend on several factors. It is understandable that if the sample size is increased the curve will become smoother (it can be presumed that the theoretical curve of diversity is smooth). This is the reason why the curves of white groups are smoother than those of the corresponding negro and ,,other races" groups. However, the number of possible categories probably also influences the fluctuations. If it is small, the fluctuations of even one frequency because of sample or age will be manifested in the curve.

There is, however, a systematic effect of the number of elements on the shape of the curves: the bias of estimates $\hat{N}_1$ and $\hat{N}_{1/3}$ decreases if the sample size is increased [17]. For this reason — especially in the case of $\hat{N}_{1/3}$ — if the sample size is small the value of estimate $\hat{N}$ depends very much on n, i.e. it increases with increasing n, approaching the real $N_{1/3}$ value. Thus, the fluctuation of n and through it the fluctuation of the frequencies may be even emphasised by the bias. The effect of statistical bias can be measured by comparing Figs. 4c and 4d. In the case of Fig. 4d it can be established that the bias may even cover the changes of the index of diversity. (As it will be seen, the division of races offers data to studying the effects of sample size - departing somewhat from our original intentions.) Because of the great effect of sample size one has to treat carefully differences like the characteristic minimum point of white males cannot be observed in negro males (Figs. 3d and 3e).

The question arises, at what sample size can one trust the conclusions drawn from the curves of N_a diversity indices. Beyond our factual examples we may refer to K e m p t o n's paper (17), although in our case the Γ-distribution of the individual frequencies cannot be presumed (I z s á k and J u h á s z - N a g y [31]) and the number of categories is also considerably smaller than the 500 assumed ones in Kempton's paper. In one of his curves (K e m p t o n [17], Fig. 3) it can be seen that 80% of the maxi-

mum possible values of $\hat{N}_{1/3}$, $\hat{N}_1$ and $\hat{N}_2$ estimates can be achived at about sample sizes of 20000, 500 and 250, respectively. Our results support these orders of magnitude, in spite of the considerably smaller number of categories.

It has to be considered also, that if the number of occuring categories is small, after the filling of the empty categories the dependence of bias on n will become smaller. It should be further noted that the dependence of $\hat{N}_a$ estimates and of N_a diversity indices in general on n is due not only to the bias: we may refer to the case of N_o ($\equiv s$) where this statement is evident. In the case of epidemiological statistics, where low frequencies are often lost because of the combinations, bias arising from the combination of low frequencies is smaller. On the basis of the curves and considerations concerning the biases it can be said as a summary that N_a diversity indices are suited to follow mortality diversity according to age if the sample size is sufficiently large. In this case, however, – as Hill already described [14] – they reflect changes in diversity essentially similarly even beside different parameters.

As far as the nowadays most widely used H' Shannon index is concerned: the great similarity of the graph of H' to the other graphs implies that the use of H' should not be rejected. But when using it, we cannot set the degree of sensitivity to small frequencies. Besides, the underestimation could be disturbing when using small number of cases. And finally the use of the $N_1 = \exp(H')$ index instead of H' is supported by the almost normal distribution of the $\hat{N}_1$ estimate, wich could be very well used from the point of view of statistics.

The statistical analysis of the curves of $s(m)$ diversity indices is much easier, because as we have mentioned already, $\hat{s}(m)$ is the MVU estimate of $s(m)$. First, we may establish that $s(m)$ curves are very similar to the curves of N_a. In certain cases the similarity is greater with the curves of N_2 and N_1 (Fig. 1b), at other times it is greater with those of $N_{1/3}$ (Fig. 1c). Perhaps it is related to the choice of parameter m. Yet, at the same time, in our experience $s(m)$ is not very sensitive to changes in m: there is no significant difference between the $s(10)$ and $s(100)$ values by age in the combined groups (Figs. 3d and 3f). It is worthwhile to note the 1974 and 1975 "other race"male total groups where $s(m)$, that is free of bias, does not follow the increase of $\hat{N}_{1/3}$ and $\hat{N}_1$ with the increase of sample size (Figs. 4a and 4b).

Because of the unambiguity of the stituation we deal only briefly with the fact that is very important for us. The comparison of the curves of the indices N and $s(m)$ with the corresponding curves of the relative (!) Brillouin index may find a great deal of similarity (Figs. 4e and 4f). This is true for almost all of the studied groups. In a previous study we have calculated the α parameter values necessary for the so called fitting to the logseries distribution [31], that can also be considered as the measure of diversity in the cases of appropriate fitting (Williams[10], Kempton – Taylor, [32]). Although the fitting results favoured the lognormal and not

the logseries model, the change of α parameter with age is surprisingly similar to changes of the relative *Brillouin* index and to those of the here studied N and $s(m)$ indices (Fig. 4e). (For the sake of completeness it has to be mentioned that beside the present ICD categories 9 further categories with small sample size were also used at the calculation of α.) The use of α parameter as an index of diversity for a distribution that is different from logseries is mentioned by others, too (Kempton — Taylor [32]). The correlation between different indices of diversity has also been described [33, 27].

Concerning the statistical behaviour of $\hat{Q}$ it is very difficult to emphasize anything beside the erratic fluctuations. At the age groups where the number of non-empty categories is small the justification of Q can be doubted. It is interesting that the fluctuations barely subside even in the cases of very great sample size, as it can be established if Figs. 3d and 3e are compared. In the cases of total groups one can see in the curves the evenly decreasing tendency of Q in the middle and upper age groups.

Considering the above from statistical point of view, we see the application of $s(m)$ index family to be the most advantageous. In the case of our investigation material, however, it has a limit in its application, for we cannot always take advantage of the fact that the sensitivity to low frequencies can be influenced by changing m: $s(m)$ is sensitive to low frequencies if m falls to the order of magnitude of the greatest n_i-s, wich condition is impossible to meet simultaneously for all age groups, because of the very different sample sizes.

Changes of diversity indices as a function of age

On the basis of the diversity curves we shall give a short description about the development of diversity as a function of age in every studied section of diseases. The description is based on the 1974 and 1975 curves of the white and negro male and female groups and on the curves of the white combined 1974 and 1975 groups (i. e. on 10 figures in each section!). Naturally, studying the great amount of figures gives further informations, too. It has to be noted, however, that the validity of the descriptions in sections III, VI, X and XIV, where there are very few categories, is quite limited. In the upper age groups (above 85) the fast changes of diversity are often not credible, because they can be traced back to statistical bias. At the description of the sections greater credit was given to the curves of groups with greater sample size.

It strikes the eye immediately in the figures that the character of the curves of indices N and $s(m)$ are mostly very similar, or if there is a difference, it can be grasped easily (comp. Hill [14], Smith et al. [27]). The behaviour of the curves of Q is basically different from and is more difficult to characterize than that of the above ones. The following description are based on indices N and $s(m)$; we will return briefly to the curves of Q at the end of the paragraph.

Diseases of the circulatory system (section VII) Diversity usually increases until the age of 20 – 24, then it decreases until the age 45 – 49, then it increases again or is stagnant. The minimum around the age of 45 – 49 is the most expressed in the white male group. $N_{1/3}$ decreases evenly after the maximum value, except in the white male group (Figs. 1a and 1b).

Neoplasms (section II) The curves remind us in many respect to those of the previous section. Diversity increases in the male group generally until the age of 30 – 34, while in the female group until the age of 20 – 24. Then it subsides until the age of 55 – 59 in the male group and until the age of 45 – 49 in the female group, then it increases again. The rise after the minimum point is more expressed in the female group (Figs. 1c and 1d). In the negro groups – where the sample size is considerably smaller – $N_{1/3}$ increases almost evenly until the upper age groups (Fig. 1d).

Infective and parasitic diseases (section I) The character of the curves differs from that of the very similar curves of sections II and VII (Figs. 1e and 1f). Diversity generally reaches a maximum value at the age of 15 – 19, then it decreases or is stagnant. In certain curves the characteristic minimum after the maximum can be recognized (Fig. 1f).

Endocrine diseases (section III) The shape of the curves is different from all of the above curves. There can be found a local maximum repeatedly around the age of 15 – 19 in the male and around the age of 20 – 24 in the female group. This is followed by a rapid decrease, finally, the diversity indices (primarily N_2 and N_1) fall to the smallest possible level (Fig. 2a).

Diseases of the nervous system (section VI) Because of the low number of categories the shape of the curves is constant. Prinarily the maximal diversity around the age of 55 – 59 can be accentuated (Fig. 2b).

Diseases of the respiratory system (section VIII) An in the beginning fast decrease after the somewhat uncertain maximum around the age of 30 can be observed in several large groups, sometimes with a shallow minimum around the age of 60 (Fig. 2c). This feature reminds us of the curves of sections II and VII.

Diseases of the digestive system (section IX) After the local maximum at the age of 15 – 19 or 20 – 24 diversity decreases steeply until the age of 45 – 49, then it increases again, often to a level surpassing even the first maximum (Figs. 2d and 2e.)

Diseases of the urogenital system (section X) After the inconstant diversity changes in the lower age groups diversity increases from the middle age group until the age of 60 – 70, or even further (Figs. 2f and 3a). In this respect there can be found a relationship with the curves of section VI that contains very few categories, too.

Congenital anomalies (section XIV) Because of the very few number of categories and low frequency of cases the shape of the curves is often very inconstant (Fig. 3b). Where the number of cases is higher a generally decreasing tendency can be accentuated (Fig. 3c).

The diversity curves of the combined material bear the features described already at several sections (Figs. 3d, 3e and 3f). It can be said again

that the hollow that often forms even a minimum is smaller in the groups of white females, negro males and females. In the case of the relative B r i l- l o u i n index it had to be emphasized that the few leading frequencies of the circulatory and neoplasmic causes of death have a decisive influence in the formation of the curves. Here, however, in the cases of the studied N_1 and $N_{1/3}$ indices it can be established that the role of middle frequencies cannot be neglected, either. This can well be illustrated by the following: if e. g. in the case of the combined 1975 negro male material the greatest frequency of 1157 of the 50 — 54 age group is decreased by 1 and a new category with the frequency of 1 is opened, in the case of N_1 diversity remains 24.8 to the nearest tenth, while in the case of $N_{1/3}$ it increases from 66.7 to 67.0, wich is a considerable change if we look at the total number of elements of 6698. (Previously P e e t [7] attempted to study the sensitivity of diversity indices (see [27], too)). In the case of "other races" we left out the curves by sections, but the curves of the combined material is presented here, too (Figs. 4a and 4b). The significant difference between the corresponding curves of the white and negro groups can be explained by the small number of elements.

The behaviour of the curves of the Q indices of diversity belonging to the section is difficult to grasp because of the irregular behaviour and because there are significant differences between the 1974 and 1975 curves of the same group. But it can be stated about the Q curves of the combined material without any doubt that the Q index of diversity increases until the age of 45 in females, then it decreases at a nearly even rate. This phenomenon deserves attention, because while in the development of N and $s\,(m)$ indices of diversity high frequencies have the leading role, in the development of Q they do not even take part. Thus, altough Q does not give a picture of the complete diversity — e. g. in the case of the combined material the first and last 40 (!) frequencies are omitted at the beginning —, it gives information about the systematic changes in the middle frequencies.

Summarizing the above, the different indices of diversity account for the development of diversity as a function of age generally in a similar way. Therefore, continuing to keep in mind their statistical characteristics, in the future we may turn our attention to the causal analysis of diversity changes as a function of age.

Sexual differences

The previously described differences [5, 6] are supported by the diversity indices studied here. The essence of the investigation was that — here, too — in every section and in every age group the diversity indices (rounded off to the nearest tenth) of males and females were compared. The event when the males' diversity index was greater was marked by the letter m, the opposite was noted by f, $\varkappa$ means that the indices were equal, and the sign — was written if the absence of one of the indices prevented the comparison. On the basis of the impressions of the patterns of the resulting tables the following conclusions can be drawn.

The three N indices and the s (m) index reflect the sexual differences very similarly, in the case of Q, however, quite different behaviour can be observed quite often. In a considerable number of the sections of diseases (sections of the diseases of the circulatory, respiratory, digestive and urogenital systems) in the first and last few age groups diversity is greater in the male group (more or less unanimously), while in the middle age groups it is greater in the female group. The dominance of male diversity in the lower age groups and that of the female diversity above the age of 30 – 40 is general. In the case of the combined material the greater male diversity of the causes of death falls back to the first and last few age groups (Table II). At the same place it can be observed that the Q index also demonstrates greater diversity in the female groups.

It would be useful in the future to investigate further the sexual differences of diversity indices on a different study material and with more categories in a section.

Table II.

Sexual differences in the White total 1974 + 1975 group.

(m and f show the greater diversity of the male and female groups, respectively)

$N_{1/3}$	m	m	f	f	f	f	f	f	f	f	f	f	f	f	f	f	x	m	m	f	
N_1	m	f	f	f	f	f	f	f	f	f	f	f	f	f	f	m	m	m	m	m	
N_2	m	f	f	f	f	f	f	f	f	f	f	f	f	f	f	m	m	m	m	m	
$s(10)$	m	f	f	f	f	f	f	f	f	f	f	f	f	f	x	m	m	m	m	m	
$s(100)$	m	–	–	f	f	f	f	f	f	f	f	f	f	f	f	m	m	m	–	–	
Q	m	m	m	f	f	f	m	m	f	f	f	f	f	f	f	f	f	m	f	m	

REFERENCES

1. H e r d a n, G. 1957. The mathematical relation between the number of diseases and the number of patients in a community', Journal of Royal Statistical Society, Ser. A, **120**, 320 – 330.
2. B o w m a n, K. O. – H u t c h e s o n, K. – O d u m, E. P. – S h e n t o n, L. R. 1971. 'Comments on the distribution of indices of diversity. In P a t i l, G. P. – P i e l o u, E. C. – W a t e r s, W. E. (Eds) *Statistical Ecology*, Vol. **3**, Pennsylvania State University Press, pp. 315 – 366.
3. E r h a r d t, K. 1977. The underutilisation of vital statistics', American Journal of Public Health, **67**, 325 – 326.
4. L i l i e n f e l d, D. E. 1977. 'On the need for general laws of mortality in epidemiology (letter)', American Journal of Epidemiology, **105**, 502 – 503.
5. I z s á k, J. – J u h á s z-N a g y, P. 1979. 'Studies on the diversity of morbidity data series', Biológia, **27**, 177 – 183. (in Hungarian, with an English summary).
6. I z s á k, J. – J u h á s z-N a g y, P. 1980. 'Investigation on diversity of Hungarian mortality statistics', Annales Univ. Sci. Budapestiensis, Sectio Biologica, **22 – 23**, (35 – 45).
7. P e e t, R. K. 1974. 'The measurement of species diversity', Annual Review of Ecology and Systematics, **5**, 285 – 307.
8. P i e l o u, E. C. 1975. *Ecological Diversity*, Wiley, New York.
9a. *Vital Statistics of the U. S. 1974, Vol. 2., Mortality.* U. S. Department of Health, Education and Welfare, Public Health Service, Govt. Print. Off., Washington, pp. 1 – 186 – 1 – 251 (1978).
9b. *Vital Statistics of the U.S. 1975, Vol. 2., Mortality.* U.S. Department of Health, Education and Welfare, Public Health Service, Hyattsville, pp. 1 – 186 – 1 – 251 (1979). 10.
10. W i l l i a m s, C. B. 1964. Patterns in the Balance of Nature, Academic Press, New York.

11. S i m p s o n, E. H. 1949. 'Measurement of diversity', Nature, 163, 688.
12. F a g e r, E. W. 1972. 'Diversity: A sampling study', American Naturalist, 106, 293 – 310.
13. H u r l b e r t, S. H. 1971. 'The nonconcept of species diversity: a critique and alternative parameters', Ecology, 52, 577 – 586.
14. H i l l, M. O. 1973, 'Diversity and evenness: a unifying notation and its consequences', Ecology, 54, 427 – 432.
15. K e m p t o n, R. A. – T a y l o r, L. R. 1976. 'Models and statistics for species diversity', Nature, 262, 818 – 820.
16. P a t i l, G. P. – T a i l l i e, C. 1979. 'An overview of diversity. In G r a s s l e, J. F. – P a t i l, G. P. – S m i t h, W. K. – T a i l l i e, C. (Eds), Ecological Diversity in Theory and Practice, International Co-operative Publishing House, Fairland, Maryland, pp. 3 – 27.
17. K e m p t o n, R. A. 1979. 'The structure of species abundance and measurement of diversity', Biometrics, 35, 307 – 321.
18. M a c A r t h u r, R. H. 1972. Geographical Ecology, Harper and Row, New York.
19. K e m p t o n, R. A. – W e d d e r b u r n, R. W. M. 1978. 'A comparison of three measures of species diversity', Biometrics, 34, 25 – 37.
20. M a c A r t h u r, R. H. 1965. 'Patterns of species diversity', Biological Reviews, 40, 510 – 533.
21. W h i t t a k e r, R. M. 1972. 'Evolution and measurement of species diversity': Taxon, 21, 213 – 251.
22. B a s h a r i n, G. P. 1959. 'On a statistical estimate for the entropy of a sequence of independent random variables' Theory of Probability and Its Applications, 4, 333 – 336.
23. S i m b e r l o f f, D. S. 1974. 'Permo-Triassic extinction: effects of area on biotic equilibrium', Journal of Geology, 82, 267 – 274.
24. S m i t h, W. – G r a s s l e, J. F. 1977. 'Sampling properties of a family of diversity measures', Biometrics, 33, 283 – 292.
25. H e c k, K. L. – V a n B e l l e, G. – S i m b e r l o f f, D. S. 1975. 'Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size', Ecology, 56, 1459 – 1461.
26. S i m b e r l o f f, D. S. 1979. 'Rarefaction as a distribution-free method of expressing and estimating diversity', In G r a s s l e, J. F. – P a t i l, G. P. – S m i t h, W. K. – T a i l l i e, C. (Eds), Ecological Diversity in Theory and Practice, International Co-operative Publishing House, Fairland, Maryland, pp. 159 – 176.
27. S m i t h, W. – G r a s s l e, J. F. – K r a v i t z, D. 1979. 'Measures of diversity with unbiased estimates' In G r a s s l e, J. F. – P a t i l, G. P. – S m i t h, W. K. – T a i l l i e, C. (Eds), Ecological Diversity in Theory and Practice, International Co-operative Publishing House, Fairland, Maryland, pp. 177 – 191.
28. E n g e n, S. 1978. Stochastic Abundance Models, Chapman and Hall, London, 1978.
29. E n g e n, S. 1979. 'Some basic concepts of ecological equitability' In G r a s s l e, J. F. – P a t i l, G. P. – S m i t h, W. K. – T a i l l i e, C. (Eds), Ecological Diversity in Theory and Practice, International Co-operative Publishing House, Fairland, Maryland, pp. 37 – 50.
30. G r a s s l e, J. F. – P a t i l, G. P. – S m i t h, W. K. and T a i l l i e, C. (Eds) 1979. Ecological Diversity in Theory and Practice, International Co-operative Publishing House, Fairland, Maryland:
31. I z s á k, J. – J u h á s z-N a g y, P. 1982. 'Studies of lognormality on mortality statistics', Biometrical Journal 24, 731 – 741.
32. K e m p t o n, R. A. – T a y l o r, L. R. 1974. 'Log-series and log-normal parameters as diversity discriminants for the lepidoptera', Journal of Animal Ecology, 43, 381 – 399.
33. A u c l a i r, A. N. and G o f f, F. G. 1971. 'Diversity relations of upland forest in the western Great Lakes area; American Naturalist, 105, 499 – 528.