INVESTIGATION ON DIVERSITY OF HUNGARIAN MORTALITY STATISTICS

by

J. IZSÁK* – P. JUHÁSZ-NAGY

Department of Plant Taxonomy and Ecology of the Eötvös Loránd University, Budapest, * College of Finance and Accountancy, Budapest

Received on 25th May, 1980

Introduction

During our previous investigations referring to hospital morbidity our attention was turned to the characteristic differences existing between male and female patient groups. Notably we found a quantity of certain entropy-like character essentially more frequently to be larger in female patient groups than in the corresponding male patient groups (5).

We notice here on the mathematical background of the question in short the following. Let us suppose a universe to be given with elements sortable in classes. The question may be raised up how evenly the individuals distribute within the classes. The distribution may be said the most even then if individuals come in equal numbers into each class, the less even obviously then if all elements come into single class while the other classes are empty. Instead of speaking on the evenness of distribution we speak often on the *diversity* of the investigational multitude referring to the given categories. In our case the investigational multitude will be the entirety of the deceaseds (or deaths) classed to the categories of diseases of the International Code of Diseases (ICD), the elements of the investigational multitude are thus the deceaseds (deaths). The classes are the categories of diseases within the sections of diseases of the A 150 List of causes of deaths. If – for example – the overwhelming majority of deaths belonging to the group of diseases investigated is classed to one single cause of death, then the diversity is insignificant, but if in turn they are distributed quite evenly within the circle of causes of deaths then the diversity is significant.

Investigations on diversity were carried out by biologists as first connected to questions of the entomology then biosociology (6, 7). Nowadays wide-spread investigations on diversity are running in connection of problems of the environmental protection (1, 2). The diversity is a concept being close relative to entropy thus we have to mention the investigations of entropy of biological systems in the historical references although the historical and intuitive background of the diversity and entropy investigations is in general entriely different.

Similar investigations in medical and altogether in human relation were not carried out but scantly (3, 4).

We accept as indicator of diversity the quantity

$$H = \frac{1}{N} \log \frac{1 \cdot 2 \cdot 3 \ldots N}{(1 \cdot 2 \cdot 3 \ldots N_1)(1 \cdot 2 \cdot 3 \ldots N_2) \ldots (1 \cdot 2 \cdot 3 \ldots N_s)},$$

were N is the number of elements of the investigational multitude, s is the number of classes, N_i is the number of elements of the 1., 2., ..., i-th class. Then in case of the given s and N the largest possible diversity is

$$H_{\max} = \frac{1}{N} \log \frac{1 \cdot 2 \cdot 3 \ldots N}{\left(1 \cdot 2 \cdot 3 \ldots \left[\frac{N}{s}\right]\right)^{s-r} \left(1 \cdot 2 \cdot 3 \ldots \left[\frac{N}{s}\right]+1\right)^{r}},$$

where $\left[\frac{N}{s}\right]$ is the integer part of the $\frac{N}{s}$ fraction, and further, r is equal to $N - s \cdot \left[\frac{N}{s}\right]$.

We investigate the relative diversity the formula of which is

$J = \frac{H}{H_{\max}}$, the meaning of which is obvious. We name J in the following *index of diversity.*

We considered the cause of the phenomenon described in our earlier publication (5) in general – in first line constitutional – factors that is why we supposed that a similar phenomenon certainly may be observed in the case of the statistics of causes of deaths too.

Investigations on diversity of mortality statistics

We selected as basis of our investigations the Hungarian statistics of causes of deaths of the years 1967 and 1975 as well which were elaborated after the VIIth respectively the VIIIth revision of the ICD according to the A 150 List of causes of deaths and can be found for example in the corresponding yearly volumes of the Hungarian Demographic Yearbook. The mortality statistics include the cases grouped in the single categories of causes of deaths breaking down by sex and age groups as well. Our selection was directed to two years distanced enough thence because we considered the investigation of stability of the results as important. The Lists of causes of deaths up-to-dated in the VIIth and the VIIIth revisions respectively differ in more than thirty items, in some places in significant extent. In spite of this the investigation of the two statistics according to our view as we describe it in the following – gives a very similar impression.

Our aim was to compare the diversity indices of the two male and female groups both classed in the complete lists and within the sections of causes of deaths.

The diversity indices were calculated by computer in such a manner that we approached the $1 \cdot 2 \cdot 3 \cdot \ldots \cdot M$ quantities by the so-called Stirling's formula.

In course of our investigations we had to face several problems arising from the nature of the material being at our disposal. One of such is the very small number of categories in certain sections of causes of deaths. In the discussion we do not may deal with the section "Symptoms, senility and ill-defined conditions" of the 1967 List and the sections "Diseases of the blood and blood-forming organs", the "Diseases of the skin and subcutaneous tissue" and the "Symptoms and ill-defined conditions" of the 1975 List which include only two-two categories of deaths. From obvious causes we do not deal furthermore with the XIth section connected to gravity and birth. On the other hand, we figure the two connected III – IV and XII – XIII section-pairs of the 1967 List of causes of deaths. With the exception of the above mentioned sections we calculated in case of every section and age group the diversity indices. We did not do it only then if the total of cases was 0 or 1 in which case the diversity index looses its meaning. We summed up the diversity indices in Tables I. and II. We indicated in these the number of categories included in the concerning section, the number of the total of cases and the rank of frequency of the heading "others" in the frequencies of the section. The last informs us somewhat how sufficient the enumeration in the section is. While stating the rank of the category "Others" we followed the principle: if the frequency of this category and that of other k categories is equally 1 then we considered the "others" category as the middle within the $k+1$ categories with frequency 1 even. All this concerns the "others" heading of 0 frequency too. For example, according to Table I we find in the category "Infective diseases" 35 causes of deaths, from age group of 0 years 22 male infants died and the number of those classed to the A 43 category "Other infective and parasitic diseases" is in this group the fifth. In some cases the choice of the heading "Others" is not unanimously definite. For example, in case of the infective diseases according to the List of the VIIth revision it is disturbing that beside the A 44 category "Others" there is separated another "Others" group too for the bacterial, viral and worm-caused diseases. We took in consideration here the traditionally last "Others" heading (A 44). But at the section of neoplasms we considered the category "Malignant neoplasms of other and unspecified sites" (after the VIIth revision the A 57 and after the VIIIth revision the A 58 group), within the section "Accidents" the category "All other accidents" (after the VIIth revision AE 147 and after the VIIIth revision AE 146) as the heading "Others". There is no "Others" heading in the section "Mental diseases" (in 1975 "Mental disorders").

Deriving from our aspect our investigations were extended also to the entire list of causes of deaths thus in the group "All diseases" fig-

Table I

Diversity indices (1967)

Section	age (years): 0		1–14	15–39	40–59	60–69	70–79	80–
I. Infective and parasitic diseases (35)	male	0.705(22,5)	0.627(32,2)	0.209(116,22)	0.182(556,24)	0.158(626,13)	0.150(494,11)	0.088(146,5)
	female	0.628(24,2)	0.526(32,2)	0.369(70,22)	0.318(174,11)	0.388(182,12)	0.285(255,11)	0.250(92,21)
II. Neoplasms (17)	male	0.558(9,2)	0.577(100,1)	0.803(362,1)	0.745(2385,3)	0.709(3969,3)	0.703(3166,3)	0.727(1065,2)
	female	04.50(13,3)	0.511(85,2)	0.842(357,1)	0.791(2498,1)	0.755(2764,1)	9.747(2943,1)	0.752(1282,1)
III. , IV. Allergic, endocrine, metabolic diseases and diseases of the blood and blood-forming organs (6)	male	0.336(5,1)	0.514(16,1)	0.556(43,1)	0.540(199,1)	0.451(446,1)	0.421(437,1)	0.323(152,1)
	female	0.000(2,2)	0.546(10,1)	0.807(39,1)	0.700(188,2)	0.542(394,2)	0.558(485,2)	0.591(175,1)
V. Mental and neurotic diseases (psychoneuroses) and pathological disorders (3)	male	–	0.154(21)	0.995(34)	0.846(55)	0.690(40)	0.546(19)	0.296(9)
	female	–	0.000(14)	0.621(19)	0.802(22)	0.490(17)	0.259(26)	0.143(23)

VI. Diseases of the nervous systems and sense organs (6)	male	0.533(78,3)	0.724(109,1)	0.834(156,3)	0.358(788,2)	0.170(1889,2)	0.080(2819,2)	0.040(1939,2)
	female	0.555(70,2)	0.799(78,1)	0.811(128,3)	0.378(704,2)	0.163(1866,2)	0.090(3881,2)	0.058(3000,2)
VII. Diseases of the circulatory system (8)	male	–	0.347(15,5)	0.403(361,7)	0.384(2801,7)	0.429(4855,5)	0.483(6141,5)	0.511(5104,5)
	female	–	0.584(23,6)	0.537(209,8)	0.559(1553,4)	0.534(3557,5)	0.544(7612,5)	0.539(7993,5)
VIII. Diseases of the respiratory system (11)	male	0.684(280,8)	0.828(81,7)	0.825(71,1)	0.689(439,1)	0.611(830,1)	0.656(889,1)	0.707(530,1)
	female	0.674(202,9)	0.753(64,5)	0.855(34,5)	0.789(166,1)	0.660(323,1)	0.692(613,1)	0.683(604,1)
IX. Diseases of the digestive system (10)	male	0.312(89,3)	0.527(44,2)	0.870(107,4)	0.769(618,4)	0.779(711,5)	0.807(564,6)	0.820(233,5)
	female	0.217(84,3)	0.573(29,2)	0.825(61,7)	0.747(368,4)	0.792(487,5)	0.781(684,5)	0.793(341,4)
X. Diseases of the genitourinary system (7)	male	–	0.481(5,5)	0.406(65,3)	0.696(159,3)	0.732(313,5)	0.622(377,5)	0.671(277,4)
	female	1.000(2,5)	0.545(9,5)	0.403(62,3)	0.651(163,4)	0.647(226,4)	0.605(263,4)	0.572(126,4)
XII., XIII. Diseases of the skin and connective tissue, diseases of the bones and of the organs of locomotion (6)	male	0.000(4,1)	0.594(7,5)	0.895(6,2)	0.898(26,4)	0.822(31,4)	0.749(19,3)	0.517(6,5)
	female	0.613(3,1)	0.613(3,4)	0.753(16,1)	0.830(36,4)	0.502(47,3)	0.523(63,2)	0.402(28,2)

Section	age (years); 0		1–14	15–39	40–59	60–69	70–79	80–
XIV. Congenital anomalies (3)	male	0.876(502,1)	0.681(73,1)	0.477(23,2)	0.000(2,2)	–	–	–
	female	0.943(396,1)	0.696(39,2)	0.577(16,2)	0.613(3,2)	–	–	–
XV. Diseases of the early infancy (6)	male	0.749(2120,2)	–	–	–	–	–	–
	female	0.746(1523,2)	–	–	–	–	–	–
XVII. Accidents, poisonings and violence (12)	male	0.381(34,1)	0.782(223,3)	0.694(1757,3)	0.624(1493,4)	0.657(663,5)	0.669(558,5)	0.541(391,5)
	female	0.521(43,1)	0.808(103,2)	0.566(391,5)	0.476(490,6)	0.592(359,5)	0.583(534,4)	0.345(608,6)
All diseases (the above not mentioned too) (138)	male	0.485(3146)	0.743(726)	0.643(3108)	0.657(9522)	0.617(14381)	0.584(15539)	0.527(10056)
	female	0.504(2362)	0.766(490)	0.753(1475)	0.684(6375)	0.615(10228)	0.560(17463)	0.499(14659)

Table II

Diversity indices (1967)

Section		age (years): 0	1–14	15–39	40–59	60–69	70–79	80–
I. Infective and parasitic diseases (35)	male	0.320(135,3)	0.432(18,20)	0.669(39,10)	0.363(277,10)	0.317(347,10)	0.312(371,13)	0.286(126,22)
	female	0.316(98,5)	0.752(9,20)	0.641(27,23)	0.509(110,25)	0.520(126,12)	0.455(180,14)	0.408(106,11)
II. Neoplasms (17)	male	0.521(16,1)	0.497(85,2)	0.727(423,1)	0.752(2569,2)	0.725(4539,2)	0.729(4796,3)	0.764(1468,1)
	female	0.545(11,3)	0.509(47,1)	0.829(399,1)	0.773(2537,1)	0.779(3244,1)	0.770(3862,1)	0.769(1976,1)
III. Endocrine and metabolic diseases (5)	male	0.328(24,1)	0.304(6,1)	0.599(11,1)	0.361(54,2)	0.356(107,2)	0.344(98,2)	0.350(17,2)
	female	0.130(17,1)	0.162(13,1)	0.543(17,1)	0.653(63,2)	0.485(210,2)	0.382(250,4)	0.263(113,3)
V. Mental disorders (3)	male	–	0.256(11)	0.885(53)	0.822(105)	0.697(67)	0.492(65)	0.231(30)
	female	–	0.000(8)	0.784(34)	0.795(48)	0.728(33)	0.294(77)	0.213(60)
VI. Diseases of the nervous system and sense organs (6)	male	0.570(117,2)	0.731(52,1)	0.645(115,2)	0.636(144,1)	0.551(172,1)	0.476(182,1)	0.260(54,1)
	female	0.545(101,3)	0.761(41,1)	0.723(78,1)	0.646(116,1)	0.527(141,1)	0.325(176,1)	0.261(132,1)
VII. Diseases of the circulatory system (9)	male	–	0.451(27,7)	0.722(519,9)	0.634(4120,8)	0.672(8013,8)	0.697(11874,8)	0.695(7786,8)
	female	–	0.443(13,6)	0.813(276,9)	0.791(2190,8)	0.738(5779,8)	0.730(13648,8)	0.716(13674,8)
VIII. Diseases of the respiratory system (8)	male	0.487(298,6)	0.798(62,5)	0.776(111,4)	0.687(485,4)	0.557(1014,3)	0.531(1546,4)	0.555(840,4)
	female	0.434(211,6)	0.776(52,8)	0.843(54,4)	0.758(199,4)	0.648(443,4)	0.654(928,4)	0.629(931,4)

IX. Diseases of the digestive system (8)	male	0.375(32,2)	0.734(19,1)	0.642(157,2)	0.601(791,3)	0.690(907,3)	0.796(799,4)	0.809(301,4)
	female	0.565(14,1)	0.550(13,1)	0.771(72,2)	0.710(422,2)	0.776(579,3)	0.817(793,4)	0.821(540,4)
X. Diseases of the genitourinary system (7)	male	0.782(4,5)	0.855(5,3)	0.512(47,3)	0.713(142,3)	0.802(309,5)	0.762(548,4)	0.735(431,4)
	female	0.564(4,1)	0.627(7,2)	0.701(55,3)	0.678(169,3)	0.631(238,3)	0.655(371,4)	0.611(241,3)
XIII. Diseases of the musculo-skeletal system and of the connective tissue (5)	male	–	0.436(4,1)	0.764(6,2)	0.861(19,3)	0.770(27,2)	0.714(30,1)	0.426(9,2)
	female	–	–	0.626(5,1)	0.700(33,2)	0.562(46,3)	0.632(81,2)	0.733(44,2)
XIV. Congenital anomalies (5)	male	0.715(603,1)	0.595(73,1)	0.590(16,2)	0.000(2,1)	–	–	–
	female	0.739(510,1)	0.575(54,1)	0.613(10,4)	–	–	–	–
XV. Some causes of perinatal morbidity and mortality (5)	male	0.833(2345,3)	–	–	–	–	–	–
	female	0.832(1740,3)	–	–	–	–	–	–
XVII. Accidents, poisonings and violence (13)	male	0.513(48,1)	0.801(195,9)	0.675(1959,7)	0.648(2214,6)	0.676(1155,5)	0.644(990,5)	0.486(512,4)
	female	0.581(37,1)	0.801(99,6)	0.576(410,9)	0.583(654,8)	0.593(557,5)	0.523(870,7)	0.336(1043,9)
All diseases (the above not mentioned too)	male	0.501(3632)	0.732(559)	0.643(3471)	0.662(10946)	0.630(16673)	0.607(21339)	0.558(11609)
	female	0.501(2748)	0.750(362)	0.764(1503)	0.700(6577)	0.646(11443)	0.588(21284)	0.538(18956)

uring in Tables I. and II. we took in consideration every category of the original list.

It is a problem too that certain enumerated causes of deaths may occur only in one or other sex. Such are the malignant nepolasms of uterus and prostata, the hyperplasms of prostata and in certain sense perhaps the malignant neoplasms of the breast. Thus in case of the sections involved we calculated the diversity indices also abandoning such categories. We disregard although the discussion of these since the conclusions below are not influenced by the modified diversity indices.

Analysis of results

The more detailed analysis of the diversity indices was carried out by comparing in every category the diversity indices of the males and females belonging to the same age group. The event that the diversity index of the male group or the female group is greater we designated by the symbols ♂ and ♀ respectively. If we could not decide because of the lack of any deversity index or the equality of the indices we designated it by "x". The Tables III. and IV. obtained thus formed the basis of the further analyses. We disregarded however in the following the diversity index of the section "Accidents" because of its nature differing from the others. Since the consideration of many pair of diversity indices is the question the enumeration of differences non-significant at the standard levels is perhaps justifiable. We loose obviously information by surpassing the quantification of the difference, but we are going to see that we get definitive answer to our supposition in this way too.

The discrepancy of Tables III. and IV. can be reduced partly to the alteration of the List A 150 of causes of deaths, partly to the random fluctuation. The systematic epidemiological changes during the past

Table III

Section number (1967)	age-group (years)						
	0	1–14	15–39	40–59	60–69	70–79	80–
I.	♂	♂	♀	♀	♀	♀	♀
II.	♂	♂	♀	♀	♀	♀	♀
III., IV.	♂	♂	♂	♀	♀	♀	♀
V.	×	♀	♂	♂	♂	♂	♂
VI.	♀	♂	♀	♂	♂	♂	♂
VII.	×	♂	♂	♂	♂	♂	♂
VIII.	♂	♂	♂	♂	♂	♂	♂
IX.	♀	♀	♂	♂	♂	♂	♂
X.	×	♀	♂	♂	♂	♂	♂
XII., XIII.	♀	♀	♀	♂	♂	♂	♂
XIV.	♂	♀	♀	♀	♂	♂	♂
XV.	♀	♀	♀	♀	×	×	×
XVII.	♀	♀	♀	♀	♂	♂	♂
All diseases	♀	♀	♀	♀	♂	♂	♂

Table IV

Section number (1975)	age-group (years)						
	0	1–14	15–39	40–59	60–69	70–79	80–
I.	♂	♀	♂♀	♀♂♀	♀♂♀♂	♀♂♀♂	♀♂♂♀♂♀
II.	♂♀	♀♂	♂♀♂	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
III.	♂	♂	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
V.	×	♂	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
VI.	♂	♂	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
VII.	×	♀	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
VIII.	♂	♂	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
IX.	♀	♂	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
X.	♂	♂	♂♀	♂♀♂♀	♂♀♂♀♂	♂♀♂♀♂	♂♀♂♀♂♀
XIII.	×	×	♀	♂	♂	×	×
XIV.	♀	♂	–	×	×	–	–
XV.	♂	–	♂♀	♂♀	–	♂	♂
XVII.	♀	×	–	×	–	×	×
All diseases	×	♀	♂♀	♀	♂♀	♂	♂

eight years play probably an unsignificant role. Comparing the values corresponding each other of the two Tables we may observe in one part of the categories the increase while in others the decrease of the diversity indices. From our point of view the difference of the events ♂, ♀ and "x" within the two years is fundamental. We summed up the result of comparison in a contingency table illustrated in Table V. The categories not agreeing in the VIIth and the VIIIth revision were naturally not taken in consideration. We could state that beside notable differences there exists a positive association between the events corresponding each other in the years 1967 and 1975.

Coming to the discussion of the results first we investigated the proportion of the total ♂ and ♀ events namely that what is the proportion of the number of pairs of groups showing greater diversity of causes of deaths in males to those pairs of groups where the diversity of causes of deaths in females is greater. The ♂:♀ proportion was in 1967 32:40, in 1975 29:41 thus we may speak in both of the cases on essential preponderance of diversity in female groups. Because of the problems men-

Table V

Contingency table of the differences of diversity indices

1967 \ 1975	♂	♀	×
♂	14	12	0
♀	6	25	1
×	1	0	5

tioned above we may consider only one part of the diversity indices as relatively confident. In selecting these we followed the principle that if the frequency of at least one out of two categories "Others" belonging to the pair of diversity indices was one of the three greatest frequency then we did not point up the ♂ or ♀ event, in the opposite case we pointed them up. As we mentioned the events of the accidents were here neglected.

Now the further investigations based already on the reduced data the proportion ♂:♀ was in 1967 10 : 14, in 1975 4 : 17 thus in both of the cases the unequivocally greater diversity of the mefale groups is observable. The proportion related to the other less convincing events was in 1967 22 : 26, in 1975 25 : 24. We cannot speak here on clear-cut sex difference. The cause of this may be for example that even in case of the distribution with greater uniformity the widening of the item "Others" leads to an extensive increase of the frequency of this category what may essentially decrease the originally great diversity.

It is noteworthy that the female preponderance of the diversity index does emerge not only in one or two sections but from the 12 analysed sections both in 1967 and 1975 as well the number of the ♀ events was greater equally in 7 – 7 sections than the number of the ♂ events.

Although the systematic deviation of the diversity index seems to be undoubtful between the two sexes the details and the validity circle of the phenomena needs further investigations which may give a basis also to the explanation of the deviation.

Summary

In spite of the problems mentioned we may take the part that in case of females the deaths are distributed in general more evenly among the categories of causes of deaths than in the case ofthe males. As the measure of the uniformity of distribution we used the relative diversity index described above. The difference between the sexes is observable within the most sections of causes of deaths.

To the analysis of the cause of the phenomenon there is a more comprehensive and detailed investigation necessary.

REFERENCES

1 Collier, B. D. et al. 1974. Dynamic Ecology. Prentice-Hall, Englewood Cliffs (N. Y.).
2 van Dobben, W. H. – Lowe-McConnell, R. H. (editors) 1957. Unifying Concepts in ecology. Junk, The Hague.
3 Herdan. G 1957. The mathematical relation between the number of diseases and the number of patients in a community. J. Roy. Stat. Soc. Ser. A., **120**. 320 – 330.
4 Höpker, W. W. 1975. Zur Quantifizierung selectionsbedingter Gruppenunterschiede in der Epidemologie. Meth. Inform. Med., **14**. 144 – 149.
5 Izsák, J. – Juhász-Nagy, P. 1979. (Analysis of morbidity data diversity.) Morbiditási adatsorok diverzitásának vizsgálata. Biológia **27**. 177 – 183.
6 Pielou, E. C. 1975. Ecological Diversity. Wiley, New York, 165 p.
7 Williams, C. B. 1964. Pattern in the Balance of Nature. Academic Press, London, 324 p.