Environmental Sciences (China), Vol. 1, No. 1, pp 65–72, 1989 ISSN 1001–0742

A study on " selenium–health " database in China

Bai Naibin[1], Zhang Keming[1] and Du Min[1]

(Received November 28, 1988)

Abstract — A database for " Se and health" has been established using DATATRIEVE data management system on a VAX–11/780 computer. Over 5000 pieces of information including 300 kinds of samples from 28 provinces in China were stored. The information relates to endemic diseases such as Keshan disease, Kaschin-Beck disease, Se–poison disease, cancers, heart disease and so on. The database can be accessed through keywords: samples, province, county, analytical method and the Se–content values can be obtained in tabulated form. The output includes primary statistics of the retrieved data sets. A multi–variable statistic program package featuring in pattern recognition is attached to the system.
Keywords: database; Selenium; endemic disease.

The study on " selenium–health " is a rather active field in the environmental sciences. A wealth of data about selenium contents in environment and humanbody have been accumulated in China. The data were gathered, put in order, then, a " selenium-health" database was established (Bai Naibin, 1988; Zhang Keming, 1988).

" SELENIUM–HEALTH" DATABASE

The data sources in the database come only from Chinese literatures. About 5000 records, 300 kinds of samples, dispersed over 28 provinces in China have now been stored for users. The units of Se-content values in samples were unified and standardized in three types: ppb (drinking water); mg/m^3 (gases); ppm (others). In general, each of sample Se–content values is average of more than 10 samples in the quoted references. The health properties contained in the database have been ranged from non–endemic to some of endemic diseases such as Keshan disease, Kaschin–Beck disease, Se–poison disease, the cancers, the endemic heart disease and others. All of data has been managed by using of DATATRIEVE data management system, and carried out in VAX–11/780 mini-computer. The database consists of three parts: management system, data files and a pattern recognition program package. The format of record in the sample data file was defined as follows:

```
01 DB.
02 ID          PIC IS 9(6), EDIT - STRING IS ZZZZZZ9.
02 PROV        PIC X(4).
02 COUNTY      PIC X(12)
02 DIS         PIC X(7).
02 SAM         PIC X(18).
02 VALUE       PIC IS 999999 \ / 999.
02 UNIT        PIC X(3).
02 METH        PIC X(4).
02 REF         PIC IS 9(3) EDIT - STRING IS ZZZ9.
```

[1]Research Center for Eco–environmental Sciences, Academia Sinica, Beijing, China

On the bases of the key words: sample (SAM), province (PROV), county (COUNTY) and analytical method (METH), the Se-content values of the samples interested can be retrieved and output in the form of table. Typical outputs are shown in Table 1 (keyword: SAM), also in Table 2 (keyword: COUNTY).

Table 1 Se-content of ancient samples from Changsha city of HN (about 2000 years ago)

ID	PROV	COUNTY	DIS	SAM	VALUE	UNIT	METH	REF
2000	HN	Changsha	—	date	0.40	ppm	NA	16
2010				date pit	0.13	ppm		16
2020				pear	0.280	ppm		16
2030				pear pit	0.026	ppm		16

Table 2 Se-content of samples in Keshan disease area of Arunqi of NM

ID	PROV	COUNTY	DIS	SAM	VALUE	UNIT	METH	REF
19040	NM	Arunqi	Keshan	potato	0.010	ppm	FA	154
19050				vegetables	0.013	ppm		154
19060				maize	0.013	ppm		154
19070				soil	0.219	ppm		154
19080				drinking water	3.700	ppb		154
19130				urine	0.002	ppm		154
19140				hair	0.095	ppm		154
19150				hair	0.141	ppm		154
19160				blood	0.021	ppm		154
19170				blood	0.029	ppm		154
19180				wheat	0.008	ppm		154
19190				maize	0.018	ppm		154
19200				sorghum	0.013	ppm		154
19210				beans	0.023	ppm		154
19220				buckwheat	0.016	ppm		154
19230				naked oats	0.010	ppm		154
19240				millet	0.012	ppm		154

With the aid of DATATRIEVE system, the database is able to automatically select the maximum or the minimum in the retrieving data sets, and calculate their averages and variances, thus the quality of single variable data in the database could have been evaluated. Besides above functions, the database also can come to a pattern recognition program package's aid. It has contained about 20 programs and has been developed in FORTRAN-77 language. Its functions are as follows: to evaluate the quality of multi-variable data in the database; to describe the disease behavior represented data in sample categories; to determine the key parameters that govern disease rules; to determine the trace element fingerprint that identifies the various samples; to develop a calssfication model that permits disease behavior indentification; to make the prediction to the class in which the sample of unknown properties must reveal.

EVALUATION ON THE DATA QUALITY IN THE DATABASE

The results, evaluated by using of the single variable statistic methods, are contained in Table 3. They show that the data in the " selenium-health " database is of better value to users. On the other hand, by using of pattern recognition methods, we have designed a classification model for evaluation of the quality of multi-variable data; the data set of samples, which consist of soil, drinking water, food, hair and blood from endemic and non-endemic areas of Keshan disease and Kaschin-Beck disease in China, have been done in the non-linear map. On the classification Fig. 1, the correct classfication ratio to 34 points of samples reached up

to 0.97, and the unusual points of samples were identified obviously. Same result was also obtained by using the linear map method. This quality evaluation show that data quality in the database is good. The pattern recognition method is one of the effective means to evaluate the mulit–dimension data quality. (Bai Naibin, 1986–1987)

Table 3 Index of data quality in the database

Area	Index	Drinking water,ppb	Soil, ppm	Maize, ppm	Wheat, ppm	Rice, ppm	Hair, ppm
	average						
	value X	0.750	0.122	0.008	0.016	0.018	0.117
	variance S	1.460	0.072	0.007	0.017	0.013	0.082
	maximu	3.700	0.275	0.041	0.078	0.070	1.000
KE	minimum	0.001	0.005	0.002	0.004	0.006	0.034
	$X+2S$ %	100	97	94	94	94	94
	$X+3S$ %	100	100	98	97	97	99
	number of samples	6	36	53	33	33	342
	average X	0.127	0.078	0.009	0.018	0.014	0.108
	variance S	0.158	0.035	0.010	0.020	0.005	0.093
	maximum	1.100	0.157	0.071	0.121	0.020	0.730
KB	minimum	0.004	0.004	0.001	0.002	0.005	0.016
	$X+2S$ %	98	93	98	98	100	95
	$X+3S$ %	100	100	98	100	100	97
	number of samples	54	15	47	45	7	187
	average X	1.404	0.234	0.036	0.056	0.064	0.305
	variance S	7.984	0.145	0.043	0.071	0.116	0.262
	maximum	83.000	0.660	0.250	0.488	0.709	1.863
NON	minimum	0.001	0.005	0.003	0.004	0.006	0.027
	$X+2S$ %	99	96	96	94	96	97
	$X+3S$ %	100	100	98	99	96	98
	number of samples	109	122	79	70	53	376

*KE —Keshan disease; KB — Kaschin–Beck disease; NON — non-endemic disease.

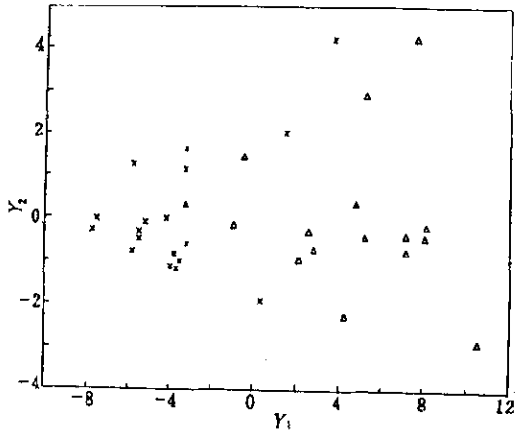

Fig.1 The non–linear map. (×—endemic area; △— non–endemic area)

A Se–CONTENT LIST FOR SOME OF POPULAR FOOD IN CHINA

At present, the study on nutriology of the life element Se gets more and more interesting because of it nutriology value. Based upon our database's data, a Se-content list for some of popular food in China has been compiled. In Tables 4, we have enumerated the Se–content to some of staple food from Keshan disease, Kaschin–Beck disease and non–endemic areas in China. Each of the data is an average value from 10 and above. The data presented in Table 4 indicate that the Se–content level of popular staple food is highly dependent upon their producing area, and Se–content levels of popular staple food produced from endemic areas are obviously low than in non–endemic areas.

Table 4 The Se–content values of some stable food from Keshan disease, Kaschin–Beck disease and non-endemic areas in China

sample	endemic, ppm	non–endemic, ppm	sample	endemic, ppm	non–endemic ppm
maize	0.010	0.035	maize flour	0.016	0.051
wheat	0.028	0.051	wheat flour	0.020	0.047
rice	0.015	0.070	beans	0.018	0.043
sorghum	0.016	0.023	soybean	0.013	0071
buckwheat	0.016	0.036	black bean	0.015	0.039
broom corn millet	0.008	0.025	kidney bean	0.010	0.115
highland barley	0.021	0.026	broad bean	0.021	0.045
naked oats	0.102	0.310	garden pea	0.012	0.055
broom corn millet	0.012	0.034	cowpea	0.008	0.027
millet	0.013	0.027	taro	0.005	0.014
oats	0.007	0.488	sweet potato	0.016	0.035
potato	0.001	0.019	dried potato	0.007	0.033

Then, a list of the Se-content value for general food in China as shown in Table 5. There is a reason for us to assume that the Se–content level of the stable food from non-endemic area, 0.02 to 0.08 ppm, is a normal capacity of nourishment capacity absorption to Chinese. According to this, 135 samples in Table 5 can be divided into three major classes: rich Se–food; general Se–food and poor Se–food. For instance, a rough trend is:

fishs, meats, birds, eggs > oil–bearing > beans > grain > vegetables > fruits.

The most richers in them are internal organs of the livestock such as heats, livers, spleens, lungs and kindeys. Conversely, the Se–content levels of fruits, vegetables, starch and sugar are much less. This suggests that the Se–content value of food is likely to be of equal rank to the protein content values contained in them. The more the protein content value in food, the more the Se–content value. Therefore, it has been proposed that the life element Selenium exists in the protein of food in the form of the amino acids. We must point out that some of the data in Table 5 is only a average value of a few samples, and only from a few non-endemic areas such as some of the typical big cities, above rule should be approximate, but it still is helpful to the related users.

A Se–CONTENT LIST FOR SOME OF THE BODY TISSUE SAMPLES IN CHINA

As is known to all, selenium is one of the essential trace elements to human body. Thus, the Se–content level contained in body tissue such as hair, blood, urine and others should be

a index to pass judgment on health level. The gathering of above mentioned data get more and more important. In order to gain a better understanding of the Se–content levels and its distribution in body tissues, the Se–content values for some of body tissue samples in our database were shown in Table 6. It is noteworthy that except hair, blood and urine, only a few samples are presented, therefore, these data in Table 6 are only for users reference.

The data presented in Table 6 indicate that: (1) the selenium–distribution in body is not well–distributed, the Se-content level of internal organs is richer; (2) the Se–content levels of all tissues in the foetus are richer than in adults; (3) the Se–content level differences between patient's and citizen's (or resident's) tissues are very clear. For example, to Keshan disease and Kaschin–Beck disease, city is higher than areas near endemic diseases, and both are higher than endemic areas. But, to cancer disease, Se–content levels of the cancer tissues are higher than the normal tissue. This encourages us to believe that the Se–content levels in some of body tissues may be a important index to judge one's health level.

Table 5 A list of the Se–content values for general food in China

Sample	Se–content, ppm	Sample	Se–content, ppm
horse kidney	4.980	highland wheat	0.049
fish flour	2.697	wheat flour	0.047
chichen liver	2.650	glutionus rice	0.046
pig kidney	2.173	chinese cabbage	0.046
cow liver	1.440	broad bean	0.045
blood clam	1.200	mutton	0.044
prawn	1.124	sweetened rice flour	0.044
loach	1.175	beef	0.043
sea crab	1.160	beans	0.043
rabbit kidney	1.040	day lily	0.041
rabbit liver	0.890	black bean	0.039
highland barley-wine	0.740	carrot	0.037
		buckwheat	0.036
entelope's horn	0.710	green soya bean	0.036
sea fish	0.648	maize	0.035
pig liver	0.589	sweet potato	0.035
mussel	0.570	green bean	0.035
fish	0.531	Chinese trumpet	
mushroom	0.530	creeper	0.035
oats	0.488	broom corn millet	0.034
frog	0.385	dried potato	0.033
goose egg	0.366	edible fungus	0.032
naked oats	0.310	anise	0.029
duck egg	0.307	barley	0.028
shrimp	0.274	green garlic leaf	0.028
rabbit heart	0.250	cowpea	0.027
pig heart	0.240	onion	0.027
chives	0.237	rapeseed	0.027
egg	0.223	highland barley	0.026
jellyfish skin	0.210	broomcorn oast	0.025
smill bean curd	0.117	dried bean curd	0.024
sunflower oil	0.176	lettuce	0.024
ciwujia	0.170	kidney bean	0.024

Table 5(continued)

sesame	0.164	cotton seed	0.023
ginseng	0.150	sorghum	0.023
peanut	0.137	lotus root	0.022
hot pepper	0.129	apple	0.020
laver	0.117	wild cabbage	0.020
fermented bean		white bean	0.019
curd	0.117	potato	0.019
kidney bean	0.115	water chestnut	0.018
tussah chrysalis	0.113	radish	0.016
pork	0.106	spinach	0.015
silkwormchrysalis	0.103	turnip bean	0.015
salted soybean	0.101	multicolour cowpea	0.014
salted vegetable	0.100	hyacinth bean	0.014
salted beancurd	0.100	taro	0.014
Korea ginseng	0.100	dried taro	0.013
sunflower seeds	0.098	walnut	0.013
kelp	0.095	black wheat	0.013
bean milk cream	0.090	yellow mustard	0.012
breast milk	0.079	mandarin orange	0.011
		green onion	0.011
substitute	0.071	ginger	0.009
bean	0.071	raisin	0.009
rice	0.070	orange	0.008
milk powder	0.069	chestnut	0.008
cabbage	0.063	starch	0.007
the root of memb-		fennel oil	0.007
ranous milk vetch	0.059	celery	0.006
garlic	0.056	garlic bolt	0.005
garden pea	0.055	sweetbell red pepper	0.003
rutabaga	0.055	sugar	0.003
prickly ash	0.054	mineral water	0.003
milk	0.053	tomato	0.003
tea	0.052	crystal sugar	0.002
red bean	0.052	hazelnut	0.002
dangshen	0.052	salt	0.002
wheat	0.051	drinking	
maize flour	0.051	water	0.001

Table 6 The Se-content values for some of body tissues samples in China

Sample	State of health	Value, ppm
hair	Keshan, Kaschin-Beck disease area	0.121
	non-disease area near endemic disease area	0.230
	non-disease area	0.287
blood	Keshan, Kaschin-Beck disease area	0.018
	non-disease area near endemic	

Table 6(continued)

	disease area	0.032
	non–disease area	0.075
urine	Keshan, Kaschin–Beck disease area	0.003
	non–disease area near endemic	
	disease area	0.011
	non–disease area	0.045
nail	citizen	0.621
toenail	resident in Keshan disease area	0.170
eye ball crystal	citizen	0.340
	cataract patient	0.650
stomach tissue	citizen	0.999
stomach cancer tissue	patient	1.085
gastric ulcer tissue	patient	1.120
lung tissue	citizen	0.420
	lung cancer patient	0.860
	citizen	0.729
	pulmonary tuberculosis patient	0.710
	lung cancer patient	0.797
lung cancer	patient	1.210
tissue	patient	1.078
muscle	foetus	1.029
tissue	resident in Kaschin–Beck disease	
	area	0.550
	resident near Kaschin–Beck	
	disease area	0.550
	citizen	1.029
	dead body in Keshan disease area	0.055
	dead body in Keshan disease area	0.037
	dead body in non–disease area	0.100
	dead body in non–disease area	0.057
heart tissue	foetus in Keshan disease area	0.385
	foetus in non–disease area	1.073
	dead body i Keshan disease area	0.040
	dead body in non–disease area	0.080
	dead body in the big city	0.140
	dead body in Keshan disease area	0.028
	dead body in non–disease area	0.032
	dead body in the big city	0.134
liver tissue	foetus	0.538
	dead body in Keshan disease area	0.040
	dead body in non–disease area	0.060
	dead body in Keshan disease area	0.080
	dead body in non–disease area	0.141
	dead body in the big city	0.250
kidney tissue	foetus in Keshan disease area	1.181
	foetus in non–disease area	2.397
milk	resident in Keshan disease area	0.003
milk powder	citizen	0.079

We believe, therefore, that the " selenium–health " database is a helpful tool to study the relationship between selenium and health.

REFERENCES

Bai Naibin et al ., The journal of Chinese Endemiology, 1988, 7(3): 142
Bai Naibin , Environmental Chemistry in Chinese, 1987, (6) 1: 78
Bai Naibin , Science Bulletin, 1987, 32 (21): 1678
Bai Naibin , Environmental Science, 1987, 8(16):14
Bai Naibin , The journal of Practical Endemiology, 1987, 2(2):95
Bai Naibin , ndemiology Bulletin, 1987, 2(1):55
Bai Naibin , The journal of Chinese Endemiology, 1986, 5(4):479
Bai Naibin , VIIIth International Conference on Computers in Chemical Research and Education, Beijing, Academia press, 1987:B-02
Zhang Kemin *et al.*, Acta Scientiae Circumstantiae, 1988, 8(4):488
Zhang Kemin *et al.*, Acta Scientiae Circumstantiae, 1989, 9(1): 42