

Application of Data Mining on the Development of a Disease Distribution Map of Screened Community Residents of Taipei County in Taiwan

Ching-Kuo Wei · Syi Su · Ming-Chin Yang

Received: 15 December 2010 / Accepted: 9 February 2011
© Springer Science+Business Media, LLC 2011

Abstract This study used data mining techniques to investigate disease forms in various administrative areas and to analyze the differences among various administrative areas in order to further draw up a disease distribution map. It is hoped that may help formulate future public health strategies and to allocate medical resources more appropriately. The major disease forms for residents under the age of 60 were hypertension, hyperglycemia and hyperlipidemia. In regard to the neighboring areas, three neighboring areas, A1, A3, and B9, shared the same disease problems with A4, A5, and B3, while two mountain-area cities, B7 and C10, experienced higher instances of liver function impairment. In terms of the clustering phenomenon among municipally graded administrative areas, the major health problems in Grade A cities were hypertension, hyperglycemia, and hyperlipidemia. The health problems such as liver function impairment and renal dysfunction were more frequently observed in Grade B and Grade C cities.

Keywords Data mining · Cluster analysis · Disease distribution map · Public health

Introduction

Taipei County is the administrative division with the largest population and area in Taiwan, with more than 3.5 million living there. Its jurisdiction includes a total of 29 administrative areas, such as the metropolitan-style cities of Banciao City, Jhonghe City, and Yunghe City, and the less populated townships of Pinghsi and Pinglin. There are also residents living in the mountains (Wulai Township) and along the seashore (Tamsui, Jinshan, Gongliao, etc.). Due to the great difference in population structure and lifestyle, it could be predicted that there should be significant differences of the disease forms in the various administrative areas. If the information on disease forms and the major health problems in each administrative area were fully obtained, it would be possible to effectively allocate resources to develop public health policies for various areas from the perspective of public health, which can enable the appropriate allocation of medical resources to meet the needs of each administrative area.

This study used data mining techniques to analyze the community-based integrated screening data from the Public Health Bureau, Taipei County in 2007. It was hoped that the use of this tool could efficiently analyze and investigate disease forms in the various administrative areas of Taipei County, as well as understand the major disease forms and the clustering effects in various administrative areas, and to further establish a health map for different age distributions in each administrative area. The information obtained could be used to assist local relevant health authorities and community-based medical institutions in promoting individualized health-related activities to community residents more effectively, as well as provide appropriate medical resources, promote beneficial health improvement projects and monitor the results of health improvement projects and the effectiveness of various health improvement activities.

C.-K. Wei
Department of Health Care Administration,
Oriental Institute of Technology,
Banciao City, Taipei County, Taiwan, R.O.C

S. Su · M.-C. Yang (✉)
Graduate Institute of Health Policy and Management,
College of Public Health, National Taiwan University,
No 17, Xu-Zhou Road,
Taipei, Taiwan, R.O.C
e-mail: mcyang637@ntu.edu.tw

Literature review

Numerous studies have shown that differences in geographical environments affect health, and that the occurrence of certain diseases is relevant to specific ethnic groups. For example, among the high-risk groups defined by the American Diabetes Association [1], Latinos, African Americans, and the residents living on Pacific islands are all ethnic groups who are more likely to suffer from diabetes. The birthplace, lifestyle, air to be breathed, environment, etc. all have an effect on the public health policies to be established [2].

Christie et al. [3] visualized the risk model and relevant incidence and mortality rates for cancer patients based on geographic information system (GIS) technology to facilitate the provision of care. GIS technology was used to analyze the data and investigate relevant cancer control policies and methods. Cutchin [4] described that over the past decade, social and environmental epidemiological studies have made important contributions by indicating the close relationship between geographical environment and health. The study gave an introduction to health geography and investigated how it revealed information on social epidemiology and health, using streets as the unit to divide areas. It found that exposure to large-scale petrochemical plants in Texas would directly and indirectly affect the environment and environmental exposure. Courtney [5] indicated that GIS is a tool beneficial to policy assessment, and the success of policies relies on the analysis on geographical location. The study investigated the distribution of nursing staff, using the GIS-based presentation of data, space and classical statistical data to establish the policies for recruiting and retaining nurses. Gregorio et al. [6] indicated that a cancer map can provide important clues, including the risk factor of geographic variations and the utilization of disease conditions and clinical services. Such geographic variations may put emphasis on on-site environmental hazards. In terms of the methods for disease testing, geographic variations can reflect the known risk factors of different prevalence rates and the changes in signals. In addition, the study suggested that geographic variations would increase the incidence of breast cancer based on the assessment of visualized representatives, disease incidence, and the potential population. Increases in the incidence rates in two or more neighboring areas may be neglected; and thus more attention should be paid to it.

The allocation of medical resources tends to be an important and confusing issue for health authorities in various countries. In order to reasonably meet the needs for community health services and to develop methods for assessing the need for community health services in the field of public medical service, Kaneko et al. [7] used GIS to develop strategies. The study investigated how to

allocate resources properly, promote interdepartmental cooperation, increase transparency, and found the need for community health services at the places where services were planned and provided. Granados et al. [8] developed an approach to perform quantitative analysis on the need for medical health. Under different circumstances, independent variables and estimates were chosen to test the alternative repression model. With the inclusion of different medical health services, the percentage of the need for medical health services in each area relatively increased with the increase in the population inside the area.

Lee et al. [9] used descriptive statistics to divide Taiwan into remote areas and non-remote areas. The analysis results indicated that subjects living in non-remote areas possess higher incomes, receive more education, obtain better medical health services and experience relatively fewer chronic diseases. On the contrary, those living in remote areas experience more chronic diseases. In addition, age and lifestyle are also important factors, which result in regional differences and income-related inequalities. Therefore, regional differences will affect the differences in socioeconomic status and psychological health status. Soares et al. [10] proposed a new approach for classification in the hope of providing information on the establishment of policies for regional development. The major indicators of the classification are the number of people, the economy, health, education, employment and culture. It divided Portugal into four areas with different levels of development, which reflects the finding about asymmetry, coastal area, and interior area. Shewchuk et al. [11] used cognitive mapping to understand the presentation of health and nursing journals. It was found that the scholars in the field of medical care management will use a clearly defined structure to provide more information, and this map can promote the committee's assessment and academic performance. Lavrac et al. [12] proposed an innovative method where data mining and visualization technology were used to support decision-making concerning public health and care in Slovenia. The purpose of using data mining and statistical methods to analyze the databases of public health institutions is to confirm the availability and accessibility of public health services in atypical areas. The results can be used to develop health authorities' health care policies.

Method

Materials and criteria for determination

This study performed an analysis on the community-based integrated screening database from the Public Health Bureau, Taipei County in 2007. The database included the data from 29 administrative areas in Taipei County. After the missing values

were excluded and the data on residents who did not experience any disease were deleted, there were a total of 22,180 effective samples. This study analyzed several common and important chronic diseases in the database. The diseases and criteria for determination are shown in Table 1. The criteria for the determination of abnormality were developed based on the values provided in clinical medical examination-related references and the disease determination criteria defined by various medical associations.

Among the 22,180 effective samples, there were a total of 8,784 male subjects (39.60%) and 13,396 female subjects (60.4%). In terms of age distribution, most subjects were aged 50–59, 2,733 males and 5,432 females, followed by 40–49, 2,712 males and 3,688 females. The statistics of the screened subjects are shown in Table 2.

Research method

This study used the two-step clustering approach. In step 1, Ward's method in hierarchical clustering was used to obtain the best cluster. Afterwards, the K-means method in non-hierarchical clustering was used to perform clustering.

Step 1: The use of Ward's Method to perform clustering

Ward's Method views each individual as a cluster first. Then each cluster was integrated with others in a specific order. The order was subject to the size of the in-group total variance after the clusters were integrated with each other. The Euclidean distance formula was used to calculate the distance between two clusters and to obtain the minimum value.

$$D_{pq} = n_p \times (\bar{x}_p - \bar{x})^2 + n_q \times (\bar{x}_q - \bar{x})^2$$

\bar{x} is the overall mean (the mean of the number of people with each disease) \bar{x}_p and n_p are the mean of G_p (the number of people with each disease) and the observed value (the number of areas occupied by each group), respectively \bar{x}_q

Table 1 Criteria for determination

Diseases	Criteria for determination
Liver function impairment	GOT(AST)≥42 or GPT(ALT)≥40
Hyperglycemia	Fasting Glucose(AC Sugar) ≥105
Hyperlipidemia	Cholesterol≥200 or TC-HDL≤35
Renal dysfunction	BUN ≥20 or Creatinine≥1.3
Hypertension	Systolic pressure ≥140 or diastolic pressure≥90

GOT(AST) Glutamic Oxaloacetic Transaminase (Aspartate aminotransferase)

GPT(ALT) Glutamic Pyruvic Transaminase (Alanine aminotransferase)

TC-HDL Total Cholesterol—High Density Lipoprotein

BUN Blood Urea Nitrogen

Table 2 Statistical table of the number of screened subjects

Age	Male (%)	Female (%)	Total (%)
40~49	2,712 (12.23%)	3,688 (16.63%)	6,400 (28.86%)
50~59	2,733 (12.32%)	5,432 (24.49%)	8,165 (36.81%)
60~69	1,930 (8.70%)	2,954 (13.32%)	4,884 (22.02%)
> 70	1,409 (6.35%)	1,322 (5.96%)	2,731 (12.31%)
Total	8,784 (39.60%)	13,396(60.40%)	22,180 (100%)

and n_q are the mean of G_q (the number of people with each disease) and the observed value (the number of areas occupied by each group), respectively

This study used a distance measure where a sample with the number of m variables was viewed as one point in the m dimensional space. In this m dimensional space, its distance was defined and the point with a shorter distance was categorized as the same type. Currently, statisticians have designed numerous calculation methods for similarity coefficients and correlation distance. This study used the widely used Euclidean Distance to measure the degree of affinity between samples.

$$d_{ij} = \left[\sum_{k=1}^m (x_{jk} - x_{jl})^2 \right]^{0.5}, (i, j = 1, 2, \dots, n)$$

d_{ij} : the distance between sample i and sample j (the Euclidean Distance between each administrative area) m : number of variables n : number of samples

Step 2: The use of K-means method to perform clustering

The mean of the cluster was used as the cluster center. There was a need to designate the K data points as the cluster centers in the beginning; however, new cluster centers would be developed in other rounds. The procedures are as follows:

- Ward's Method was used to decide the number of clusters and to choose the centers of clusters.
- The distance between each data and every center was calculated to compare which point was closer to it and to view that point as in the cluster.
- The mean of data in each cluster was calculated to obtain the new center.
- The center obtained was used to re-cluster each data.

The clustering can be terminated until the cluster of each data was fixed and cannot be divided into other clusters. Clustering Dendrogram is shown in Fig. 1.

This study used the cluster analysis of data mining to analyze the disease forms and the clustering effect in the 29 administrative areas of Taipei County. In our statistical analyses were performed using the Statistica 8 package. For example, the disease forms in Taipei County were divided into five clusters. In cluster 1, 50% of the residents experienced hyperlipidemia levels and the proportion was the highest in

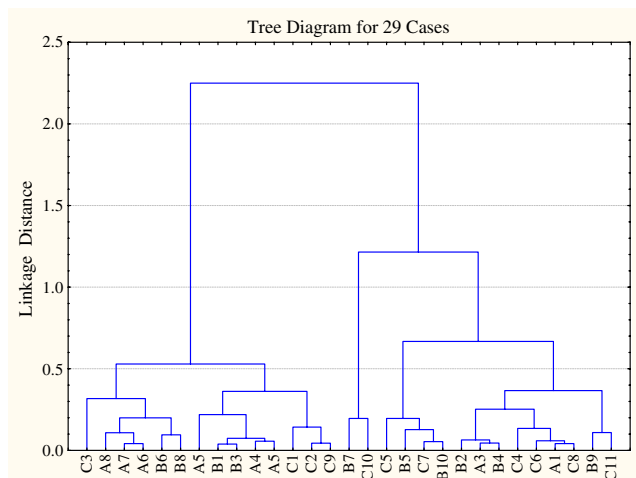


Fig. 1 Clustering dendrogram

these five clusters, suggesting that the most important disease was hyperlipidemia. The distribution of the administrative areas in these five clusters were then observed to draw develop a disease distribution map based on the regional distribution.

Results

In this study, age distribution and gender were discussed respectively. The subjects were divided into male/female at the age of 40–49, 50–59, 60–69, and over the age of 70. As the level of development in each city/township in Taipei County is different, the municipal grading system used for urban and rural development projects in Taipei County was used to divide the cities and townships into three grades: Grade A for highly developed areas, including a total of eight cities, with A1 ~ A8 used to represent them; Grade B for moderately developed areas, including a total of ten townships, with B1 ~ B10 used to represent them; Grade C for the least developed areas, including a total of 11 townships, with C1 ~ C11 used to represent them. Details are shown in [Appendix A](#).

The analysis on the diseases of male subjects at different ages is shown in [Table 3](#).

As shown in [Table 3](#), the highest proportions of the diseases experienced by male subjects aged 40–49 were hyperglycemia (28.06%), followed by hyperlipidemia in Cluster 4 (26.44%), and hypertension in Cluster 1 (25.93%). In terms of municipal grading, it was found that approximately half of the Grade A cities were in the hyperlipidemia cluster, suggesting that the primary health problem of male subjects aged 40–49 had hyperlipidemia. As for the male subjects aged 50–59, the highest proportions of the diseases experienced by them were hyperlipidemia in Cluster 2 (30.55%), followed by hyperglycemia (25.28%) in Cluster 1 and hypertension in Cluster 4 (23.56%). In this age group, higher incidences of hyperlip-

idemia, hypertension, and hyperglycemia were the major factors affecting health. According to the examinations on the male subjects aged 60–69, the highest proportion of the diseases experienced by them were renal dysfunction in Cluster 2 (38.91%), followed by hypertension in Cluster 1 (36.01%) and hyperlipidemia in Cluster 3 (25.08%). In this age group, most of the patients experienced hypertension and renal dysfunction; and it was found that the proportion of the subjects with renal dysfunction was quite high. In addition, most patients lived in Grade B (80%) and Grade C (55%) cities. The clustering of renal dysfunction in non-major urban cities was extremely significant. For the male subjects over the age of 70, according to the examination results, the highest proportion of the diseases experienced by them was renal dysfunction in Cluster 2 (42.44%). In addition, more than half of the subjects in each differently-graded cities were in this Cluster (A: 50%, B: 60%, C: 63.6%). As opposed to other age groups, the proportion of renal dysfunction in male subjects over the age of 70 was very high.

The clustering effect in female subjects and the result of sample clustering is shown in [Table 4](#).

In Taipei County, the highest proportion of diseases experienced by female subjects aged 40–49 was hyperlipidemia (40.89%), followed by hyperglycemia in Cluster 2 (35.93%), and hypertension in Cluster 1 (23.18%). For the disease distribution of the residents in this age group, most of them experienced hyperlipidemia and hyperglycemia, and only the residents in Grade C areas (C1, C6, and C9) experienced significant hypertension. The three clusters of the subjects aged 50–59 were hypertension, hyperglycemia, and hyperlipidemia. For subjects aged 60–69, the highest proportion of the diseases experienced by them were hyperlipidemia levels in Cluster 1 (40.68%) followed by hyperglycemia in Cluster 2 (40.49%). Except for the B4 and C1 areas, almost all the subjects in each area suffered from hyperlipidemia and hyperglycemia. Therefore, hypertension, hyperglycemia, and hyperlipidemia were major disease forms for aged 40–69 group. As for subjects over the age of 70, the highest proportion of the diseases experienced by them were renal dysfunction in Cluster 2 (35.40%), followed by hypertension in Cluster 3 (26.48%), and hyperglycemia in Cluster 4 (23.75%). The results were consistent with that of the male subjects. As opposed to the subjects in other age groups, the proportion of renal dysfunction in this age group was very high.

The analysis results of the overall male and female subjects in Taipei County are shown in [Table 5](#).

The results of disease clustering of the overall subjects in Taipei County are shown in the [Table](#) above. A regional distribution map of disease forms of the 29 administrative areas in Taipei County was developed based on the municipal grading and disease-clustering phenomenon, as shown in [Fig. 2](#).

Table 3 The results of clustering of diseases in male subjects

Age	Clusters	Abnormalities	Areas	Samples	%
40~49	1	Hypertension	A2, B2, C1, C7, C9, C10	685	25.26%
	2	Hyperglycemia	A3, A4, B1, B3, B4, B6, B8, B10, C2, C3, C4, C5, C11	761	28.06%
	3	Liver function impairment	A6, B7, C6	549	20.24%
	4	Hyperlipidemia	A1, A5, A7, A8, B5, B9, C8	717	26.44%
50~59	1	Hyperglycemia	A2, A3, A6, B3, B4, B7, C2, C10	699	25.28%
	2	Hyperlipidemia	A5, A8, B1, B2, B5, B8, B9, B10, C4, C5, C6, C8, C10, C11	835	30.55%
	3	Renal dysfunction	A1, B6, C7	555	20.31%
	4	Hypertension	A4, A7, C1, C3	644	23.56%
60~69	1	Hypertension	A1, A4, A6, A7, A8, B7, B10, C2, C3, C5, C10	695	36.01%
	2	Renal dysfunction	A5, B1, B2, B3, B4, B5, B6, B8, B9, C1, C4, C6, C7, C9, C11	751	38.91%
	3	Hyperlipidemia	A2, A3, C8	484	25.08%
>70	1	Hyperglycemia	B3, B7, C3	303	21.05%
	2	Renal dysfunction	A3, A4, A5, A6, B1, B2, B4, B6, B9, B10, C1, C2, C5, C7, C6, C8, C11	598	42.44%
	3	Hypertension	A1, A2, A7, A8, B5, B8, C4, C9	508	36.05%

As shown in the Figure 2 and Table 5, there were five clusters of disease forms. The highest proportion of health abnormalities experienced by the overall residents screened in Taipei County was hyperlipidemia (31.9%). The disease was distributed in Grade A, B, and C cities and townships, suggesting that hyperlipidemia was the most common overall health problem for the subjects in Taipei County, followed by hypertension and hyperglycemia, which were common in these cities and townships as well. The cumulative proportion of these three diseases was 88.9%, revealing that hypertension, hyperglycemia, and hyperlipidemia were the most important factor affecting public health. It was interesting to note that the distribution of these three diseases in Grade A cities and Grade B cities were more even. In Grade C cities, the proportion of hyperlipidemia

was higher (in 4 clusters, 36.4%), while that of hyperglycemia was the lowest (in 1 cluster, 9%). As for renal dysfunction and liver function impairment, their distribution was mainly in Grade B cities and Grade C cities. This was important information for health education promotion.

Discussion

The occurrence of diseases is highly correlated with the race, age and lifestyle of the people. Ford [13] found in a survey that the adult population of the US has a common metabolic syndrome, while race, age, obesity, and lifestyle, including smoking and physical activity, affect the prevalence of metabolic syndrome. Analyses of the geographic

Table 4 The results of clustering of diseases of female subjects

Age	Clusters	Abnormalities	Areas	Samples	%
40~49	1	Hypertension	C1, C6, C9	855	23.18%
	2	Hyperglycemia	A4, A5, B3, B5, B8, C2, C3, C8, C10, C11	1,325	35.93%
	3	Hyperlipidemia	A1, A2, A3, A6, A7, B1, B2, B4, B6, B7, B9, B10, C4, C5, C7	1,508	40.89%
50~59	1	Hypertension	A4, A5, B1, B7, C2, C7, C9	1,572	28.94%
	2	Hyperglycemia	A1, A8, B3, B4, B6, B8, C1, C3, C4, C5, C6, C10, C11	1,936	35.64%
	3	Hyperlipidemia	A2, A3, A6, A7, B2, B5, B9, B10, C8	1,924	35.42%
60~69	1	Hyperlipidemia	A2, A3, A4, A7, B2, B6, B7, B8, B9, C3, C6, C7, C9	1,207	40.86%
	2	Hyperglycemia	A1, A5, A6, A8, B1, B3, B5, B10, C2, C4, C5, C8, C10, C11	1,196	40.49%
	3	Hypertension	B4, C1	551	18.65%
>70	1	Hyperlipidemia	B7, C3	190	14.37%
	2	Renal dysfunction	A1, A3, A4, A5, A6, B1, B3, B4, B5, B9, C1, C4, C7, C8, C11	468	35.40%
	3	Hypertension	A2, A8, B2, B8, C2	350	26.48%
	4	Hyperglycemia	A7, B6, B10, C5, C9	314	23.75%

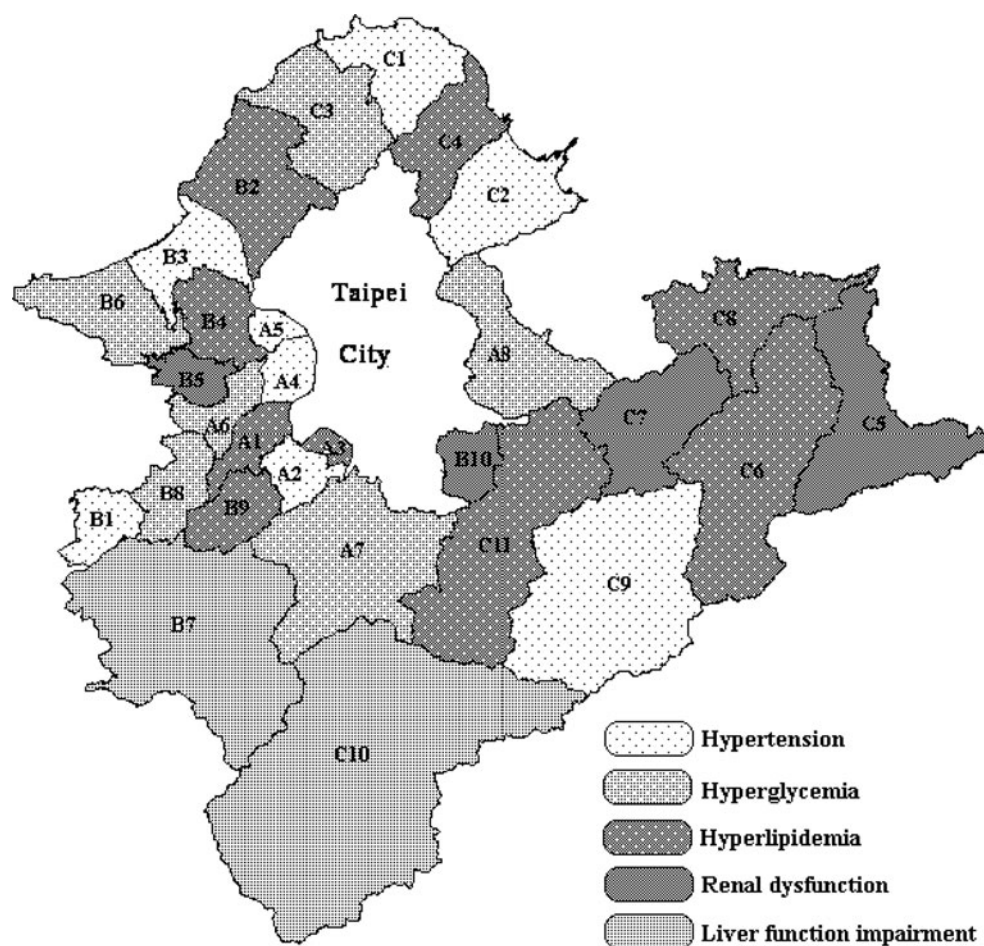
Table 5 Diagram of the disease distribution of the overall subjects in Taipei County

Clusters	Abnormalities	Areas	Samples	%
1	Hyperlipidemia	A1, A3, B2, B4, B9, C4, C6, C8, C11	7,067	31.9%
2	Hyperglycemia	A6, A7, A8, B6, B8, C3	6,412	28.9%
3	Hypertension	A2, A4, A5, B1, B3, C1, C2, C9	6,240	28.1%
4	Renal dysfunction	B5, B10, C5, C7	1,623	7.3%
5	Liver function impairment	B7, C10	838	3.8%

regions and health information can improve the disease prevention and health plans, and effectively control the increasing medical costs [14, 15].

It can be inferred from the analysis of gender and age that different disease forms and health problems are experienced by residents of different ages. The major factors affecting the health of male residents aged 40–59 were hypertension, hyperglycemia, and hyperlipidemia. For subjects over the age of 60, renal dysfunction was a common health problem. The major disease affecting female residents' health was hyperlipidemia, while renal dysfunction was also an important factor affecting the health of female residents over the age of 70. In general, hypertension, hyperlipidemia and hyperglycemia were major disease forms experienced by the general public

under the age of 60, with hyperlipidemia of particular importance. Therefore, in regard to health promotion, correct dietary and lifestyle habits of the general public should be enhanced to reduce the intake of high cholesterol food such as egg yolk, animal organs and seafood. In addition, developing a weekly exercise habit is the best way to increase health. Renal dysfunction starts to occur at the age of 60 and becomes quite significant in the general public over the age of 70, which reveals the unique disease distribution phenomenon in Taiwan. The proportion of the population receiving dialysis in Taiwan is the highest in the world. The traditional medication habits of the general public, especially the elderly, have a great influence on renal function. The elderly tend to view drugs as tonics, therefore, it is necessary to promote accurate concepts in the future to reduce such phenomenon.

Fig. 2 Disease distribution map of the overall subjects in various administrative areas of Taipei county

In addition, in terms of the clustering phenomenon among various municipally graded administrative areas, it is obvious that the major problem for Grade A cities was hypertension, hyperglycemia, and hyperlipidemia, which were also common problems in highly developed cities. Diseases such as liver function impairment and renal dysfunction occurred more frequently in Grade B and Grade C cities. In terms of geographical location, there was no significant difference in the disease distribution in the areas along the seashore. Moreover, the neighboring areas, A1, A3, and B9 shared common disease problems, as do the neighboring areas, A4, A5, and B5. However, as for liver function impairment in B7 and C10, both of them were mountain-area townships, and the major population in C10 aboriginal; whose dietary habits are different from that of the general Han population. The aspect of ethnic differences may be further discussed in the future.

Acknowledgements The community-based integrated screening data provided by the Public Health Bureau, Taipei County Government, and the research funding sponsored by Taipei County Hospital, which assisted in the conduct of the analyses in this study, are highly appreciated.

Appendix A

Table 6 Taipei County cities and townships classification

A Highly developed areas	B Moderately developed areas	C The least developed areas
A1—Banciao	B1—Yingge	C1—Shihmen
A2—Jhonghe	B2—Tamsui	C2—Wanli
A3—Yongho	B3—Bail	C3—Sanjhih
A4—Sanchong	B4—Wugu	C4—Jinshan
A5—Lujhou	B5—Taishan	C5—Gongliao
A6—Sinjhuang	B6—Linkou	C6—Shuangsi
A7—Sinclian	B7—Sansia	C7—Pingshi
A8—Sijhih	B8—Shulin	C8—Rueifang
	B9—Tucheng	C9—Pinglin
	B10—Shenkeng	C10—Wulai
		C11—Shihding

References

1. American Diabetes Association, Standards of medical care for patients with diabetes mellitus. *Diab. Care* 25(1):33–49, 2002.
2. Dummer, T. J., Health geography: supporting public health policy and planning. *CMAJ* 178(9):2008, 2008.
3. Ghetian, C. B., Parrott, R., Volkman, J. E., and Lengerich, E. J., Cancer registry policies in the United States and geographic information systems applications in comprehensive cancer control. *Health Policy* 87(2):185–193, 2008.
4. Cutchin, M. P., The need for the new health geography in epidemiologic studies of environment and health. *Health Place* 13(3):725–742, 2007.
5. Courtney, K. L., Visualizing nursing workforce distribution: policy evaluation using geographic information systems. *Int. J. Med. Inform.* 74(11):980–988, 2005.
6. Gregorio, D. I., and Samociuk, H., Breast cancer surveillance using gridded population units, Connecticut, 1992 to 1995. *Ann. Epidemiol.* 13(1):42–49, 2003.
7. Kaneko, Y., Takano, T., and Nakamura, K., Visual localisation of community health needs to rational decision-making in public health services. *Health Place* 9(3):241–251, 2003.
8. Granados, R. M., Aguilera, J. D., and Martin, J. J., Estimation of an index of regional health needs in Spain using count regression models with filter. *Health Policy* 81(1):4–16, 2007.
9. Lee, M. C., and Jones, A. M., Understanding differences in income-related health inequality between geographic regions in Taiwan using the SF-36. *Health Policy* 83(2):186–195, 2007.
10. Soares, J. O., Marques, M. M. L., and Monteiro, C. M. F., A multivariate methodology to uncover regional disparities: a contribution to improve European Union and governmental decision. *Eur. J. Oper. Res.* 145:121–135, 2003.
11. Shewchuk, R. M., O'Connor, S. J., and Williams, E. S., Using cognitive mapping to understand what health care journals represent. *Soc. Sci. Med.* 62:1192–1204, 2006.
12. Lavrac, N., Bohanec, M., Pur, A., Cestnik, B., Debeljak, M., and Kobler, A., Data mining and visualization for decision support and modeling of public health-care resources. *J. Biomed. Inform.* 40:438–447, 2007.
13. Ford, E. S., Giles, W. H., and Dietz, W. H., Prevalence of the metabolic syndrome among US adults findings from the Third National Health and Nutrition Examination Survey. *JAMA* 287(3):356–359, 2002.
14. Zhan, F. B., Brender, J. D., Han, Y., Suarez, L., and Langlois, P. H., GIS-EpiLink: a spatial search tool for linking environmental and health data. *J. Med. Syst.* 30(5):405–412, 2006.
15. Yu, W., Cowper, D., Berger, M., Kuebler, M., Kubal, J., and Manheim, L., Using GIS to profile health-care costs of VA quality-enhancement research initiative diseases. *J. Med. Syst.* 28(3):271–285, 2004.