Spatial distribution patterns of anorectal atresia/stenosis in China: Use of two-dimensional graph-theoretical clustering

Ping Yuan, Liang Qiao, Li Dai, Yan-Ping Wang, Guang-Xuan Zhou, Ying Han, Xiao-Xia Liu, Xun Zhang, Yi Cao, Juan Liang, Jun Zhu

Abstract

AIM: To investigate the spatial distribution patterns of anorectal atresia/stenosis in China.

METHODS: Data were collected from the Chinese Birth Defects Monitoring Network (CBDMN), a hospital-based congenital malformations registry system. All fetuses more than 28 wk of gestation and neonates up to 7 d of age in hospitals within the monitoring sites of the CBDMN were monitored from 2001 to 2005. Two-dimensional graph-theoretical clustering was used to divide monitoring sites of the CBDMN into different clusters according to the average incidences of anorectal atresia/stenosis in the different monitoring sites.

RESULTS: The overall average incidence of anorectal atresia/stenosis in China was 3.17 per 10,000 from 2001 to 2005. The areas with the highest average incidences of anorectal atresia/stenosis were almost always focused in Eastern China. The monitoring sites were grouped into 6 clusters of areas. Cluster 1 comprised the monitoring sites in Heilongjiang Province, Jilin Province, and Liaoning Province; Cluster 2 was composed of those in Fujian Province, Guangdong Province, Hainan Province, Guangxi Zhuang Autonomous Region, south Hunan Province, and south Jiangxi Province; Cluster 3 consisted of those in Beijing Municipal City, Tianjin Municipal City, Hebei Province, Shandong Province, north Jiangsu Province, and north Anhui Province; Cluster 4 was made up of those in Zhejiang Province, Shanghai Municipal City, south Anhui Province, south Jiangsu Province, north Hunan Province, north Jiangxi Province, Henan Province, Shanxi Province and Inner Mongolia Autonomous Region; Cluster 5 consisted of those in Ningxia Hui Autonomous Region, Gansu Province and Qinghai Province; and Cluster 6 included those in Shaanxi Province, Sichuan Province, Chongqing Municipal City, Yunnan Province, Guizhou Province, Xinjiang Uygur Autonomous Province and Tibet Autonomous Region.

CONCLUSION: The findings in this research allow the display of the spatial distribution patterns of anorectal atresia/stenosis in China. These will have important guiding significance for further analysis of relevant environmental factors regarding anorectal atresia/stenosis and for achieving regional monitoring for anorectal atresia/stenosis.

Key words: Spatial distribution; Anorectal atresia/stenosis; Two-dimensional graph-theoretical clustering; Incidence; Monitoring

Peer reviewer: Dr. Stephen E Roberts, Senior Lecturer in Epidemiology, School of Medicine, Swansea University, Honorary Research Fellow, Department of Public Health, University of Oxford, School of Medicine, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom

Yuan P, Qiao L, Dai L, Wang YP, Zhou GX, Han Y, Liu XX, Zhang X, Cao Y, Liang J, Zhu J. Spatial distribution patterns of anorectal atresia/stenosis in China: Use of two-dimensional graph-theoretical clustering. World J Gastroenterol 2009; 15(22): 2787-2793 Available from: URL: http://www.wjgnet.com/1007-9327/15/2787.asp DOI: http://dx.doi.org/10.3748/wjg.15.2787
INTRODUCTION

Anorectal atresia/stenosis is a congenital malformation characterized by absence of continuity of the anorectal canal or of communication between rectum and anus, or narrowing of anal canal, with or without fistula, to neighboring organs[1]. Its incidence is rated high amongst gastrointestinal tract malformations. Incidence relates not only to genetic factors but also to environmental factors, especially spatial differences. There is, however, very little information available in literature about the spatial distribution patterns of anorectal atresia/stenosis in China.

Since 1986 China has been using the hospital-based Chinese Birth Defects Monitoring Network (CBDMN) to dynamically monitor severe congenital malformations such as anorectal atresia/stenosis[1]. We conducted this research to divide monitoring sites of the CBDMN into different clusters using two-dimensional graph-theoretical clustering analysis of the incidences of anorectal atresia/stenosis. Consideration was given to the similarities of the incidences of anorectal atresia/stenosis and the adjacent spatial relationships among different monitoring sites. This paper will present the spatial distribution patterns of anorectal atresia/stenosis and hopes to provide clues for research on its etiology.

MATERIALS AND METHODS

Objects
Research subjects were all perinatal fetal births more than 28 wk of gestation and neonates up to 7 d of age monitored in hospitals in the monitoring sites of the CBDMN from 2001 to 2005. They included live births, fetal deaths, stillbirths and those neonates who died within the first 7 d in these hospitals.

Monitoring hospitals
Using the hospital-based guidelines for monitoring birth defects in developing countries, as recommended by World Health Organization (WHO), the CBDMN gathered data from about 460 hospitals in this hospital-based network. These hospitals - all of them above the county level - were located in 138 cities (138 monitoring sites) of 31 different provinces, municipal cities, and autonomous regions in China. The selection of monitoring sites used the method of stratified sampling based on the combination of geographical location, economic development level and infant mortality rate. The spatial distribution of monitoring sites is in accordance with the distribution of nationwide births. The nationwide program covers approximately 450,000-500,000 births annually through all monitoring hospitals.

Information collection
The monitoring staff all received technical training on the case ascertainment of birth defects and the reporting of register forms. The monitoring hospitals collected the basic monthly information about the fetuses and neonates from units of delivery, pediatric and pathology quarterly reports and filled in the “Quarterly Form for Perinatal Births”. The monitoring staff in these hospitals filled in the “Registration Card for Births with Congenital Malformations” regarding the cases of diagnosed anorectal atresia/stenosis. All the forms were required to be handed over to the provincial birth defects monitoring offices; these would be reported to the National Center for Birth Defects Monitoring after scrutiny. The specific monitoring methods and quality control measures complied with those in reference[3].

Inclusion and exclusion criteria
The perinatal births diagnosed as having anorectal atresia/stenosis with reference to criteria in Code Q42.1 and Code 42.3 in ICD-10 were included in this research. According to the criteria authorized by the International Clearing house for Birth Defects Surveillance and Research (ICBDSR), cases of mild stenosis which did not need correction and ectopic anus were excluded.

Spatial distribution analysis
The Excel Package was used to build the database of data of anorectal atresia/stenosis by monitoring sites. The ArcView GIS 3.2 was applied to spatially display the average incidences of anorectal atresia/stenosis in different provinces, municipal cities, and autonomous regions.

Two-dimensional graph-theoretical clustering:
The graph is a set of vertices and edges that connect pairs of vertices in the space[4,5]. According to the basic requirements for clustering the two-dimensional ordinal samples, the similarities of the disease-related variables between members of the same cluster and their disparities between members of different clusters need to be maintained. The connectivity of the geographic units within the cluster also needs to be preserved. The weighted connected graph was supposed to be $G = (V, E, D)$, in which (1) $V$ represents the set of the locations of the geographic units (referred to monitoring sites in this research), (2) $E$ represents the initial location connection matrix $B[0]$ (Formula 1), the set of the adjacent relationships among different monitoring sites, and (3) $D$ represents the initial disease-related distance matrix $D[0]$ (Formula 2), the weights between different vertices in the tree algorithm in the graph theory. Based on the weighted connected graph $G = (V, E, D)$, minimum spanning trees (MST) which were of biogeographic significance[6,7] were constructed by the Kruskal MST algorithm[8]. The two vertices with the minimum distance measures were selected and connected. One of the remaining vertices was selected and connected with the one of the two connected vertices to which it showed the minimum distance measured. The other remaining vertices were connected consecutively with those vertices already connected in the same manner until all the vertices were interconnected. The whole process was completed by the DPS7.05 software package[9].
\[ B^{(0)} = (b^{(0)}_{ik})_{n \times n}; \quad i, k = 1, 2, \ldots, n \quad \text{(Formula 1)} \]
\[ D^{(0)} = (d^{(0)}_{ik})_{n \times n}; \quad i, k = 1, 2, \ldots, n \quad \text{(Formula 2)} \]

Where \( b^{(0)}_{ik} \) is the labeling of location connection between the monitoring site \( i \) and the monitoring site \( k \). The value of \( b^{(0)}_{ik} \) is 1 if the two monitoring sites are adjacent, while it is 0 when the two monitoring sites are not. \( d^{(0)}_{ik} \) is the similarity distance between the incidence of the monitoring site \( i \) and that of the monitoring site \( k \).

The MST was deconstructed by the method of “necks” in the graph theory\(^{11}\). Specific steps processed were as follows: (1) calculating the “branches”: All \( n \) vertices were interconnected by (\( n-1 \)) edges. Two of these vertices were connected only by one edge and the others were connected by at least two edges, which thus formed a chain without circuits, called the “branch”. The branch with the most edges was called the main branch (or diameter) of the MST; (2) calculating the “subsidiary main branch”. Starting from any vertex in the main branch of MST, the branch with the most edges, other than the main branch, was separated out and called the subsidiary main branch. The number of edges of the subsidiary main branch was called the “depth” of the vertex; (3) identifying “necks”. The task was twofold: (1) to appoint an integer “a” (that is \( > 1 \)) and (2) to find the subsidiary main branch of every vertex with a depth \( \geq a \) in the main branch. The edges, which connected the vertices with the depth of 0 in the main shared parts of every subsidiary main branch, were called the “neck”; and (4) the necks were deleted in the graph to deconstruct the MST into parts so that the monitoring sites in the graph were divided into different clusters accordingly.

The layer of the deconstructed MST was added to the Administrative Boundary Layer of the 1:4M-scale Topographic Database of the National Fundamental Geographic Information System of China to formulate the two-dimensional MST-based cluster graph. The clustering results were used to make another cluster map for visual observation. This process was performed by the ArcView GIS 3.2 package software.

## RESULTS

A total of 2,670,367 perinatal births were monitored from 2001 to 2005 all over China. Eight hundred and forty six cases of anorectal atresia/stenosis were found, equating to a total average incidence of 3.17 per 10,000. See Table 1 for the average incidences of anorectal atresia/stenosis in different provinces, municipal cities or autonomous regions. The top five incidences appeared in Liaoning (4.89 per 10,000 births), Zhejiang (4.83 per 10,000 births), Guangdong (4.78 per 10,000 births), Chongqing (4.59 per 10,000 births) and Beijing (4.10 per 10,000 births).

### Spatial distribution

Regarding the geographic division standard for Eastern, Middle and Western China from the National Bureau of Statistics of China in 2003\(^{12}\), the areas (provinces, autonomous regions, or municipal cities) with the highest average incidences of anorectal atresia/stenosis were concentrated in Eastern China, while the areas with the lowest average incidence of less than 1.99 per 10,000 were mostly located in Western China (Figure 1).

### Results of the two-dimensional graph-theoretical clustering

The MST was constructed with consideration to the similarities of average incidences of anorectal atresia/stenosis and the spatial connectivity between different monitoring sites (Figure 2). According to the “neck” calculation method in the graph theory, when the integral constant, a, was designated as 2, the monitoring sites were divided into 6 clusters of different areas.

Regarding the average incidences of anorectal atresia/stenosis in different provinces, municipal cities and autonomous regions from 2001 to 2005, the monitoring sites were grouped into 6 clusters. Cluster 1 comprised the monitoring sites in Heilongjiang Province, Jilin Province, and Liaoning Province; Cluster 2 was composed of those in Fujian Province, Guangdong Province, Hainan Province, Guangxi Zhuang Autonomous Region, south Hunan Province, and so on.

### Table 1 Average incidences of anorectal atresia/stenosis in different provinces, municipal cities or autonomous regions of China from 2001 to 2005

| Province/Autonomous region/Municipal city | Perinatal births | Cases | Average incidence (per 10,000) |
|------------------------------------------|------------------|-------|--------------------------------|
| Liaoning                                 | 79760            | 39    | 4.89                           |
| Zhejiang                                 | 111,690          | 54    | 4.63                           |
| Guangdong                                | 127,648          | 61    | 4.78                           |
| Chongqing                                | 60,979           | 28    | 4.59                           |
| Beijing                                  | 114,741          | 47    | 4.10                           |
| Guangxi                                  | 64,444           | 26    | 4.03                           |
| Tianjin                                  | 62,542           | 25    | 4.00                           |
| Anhui                                    | 100,631          | 36    | 3.58                           |
| Fujian                                   | 95,970           | 34    | 3.54                           |
| Ningxia                                  | 65,064           | 23    | 3.53                           |
| Jiangsu                                  | 149,984          | 52    | 3.47                           |
| Jilin                                    | 112,606          | 38    | 3.37                           |
| Shanxi                                   | 76,653           | 24    | 3.13                           |
| Henan                                    | 126,082          | 39    | 3.09                           |
| Hebei                                    | 59,964           | 18    | 3.00                           |
| Hebei                                    | 130,441          | 39    | 2.99                           |
| Gansu                                    | 61,629           | 18    | 2.92                           |
| Hunan                                    | 85,919           | 25    | 2.91                           |
| Heilongjiang                             | 71,224           | 20    | 2.81                           |
| Hainan                                   | 54,234           | 14    | 2.58                           |
| Shandong                                 | 164,141          | 42    | 2.56                           |
| Guizhou                                  | 58,695           | 15    | 2.56                           |
| Sichuan                                  | 88,783           | 22    | 2.48                           |
| Shanghai                                 | 144,361          | 34    | 2.36                           |
| Inner Mongolia                           | 63,061           | 14    | 2.22                           |
| Yunnan                                   | 85,598           | 17    | 1.99                           |
| Jiangxi                                  | 84,015           | 15    | 1.79                           |
| Xinjiang                                 | 56,214           | 10    | 1.78                           |
| Shaanxi                                  | 64,634           | 11    | 1.70                           |
| Qinghai                                  | 36,666           | 6     | 1.64                           |
| Tibet                                    | 12,194           | 0     | 0.00                           |
| **Total**                                | **2,670,367**    | **846** | **3.17** |

www.wjgnet.com
Cluster 1 consisted of those in Guizhou Province, north Jiangxi Province; Cluster 2 consisted of those in Sichuan Province, Chongqing, southwest and central Jiangxi Province; Cluster 3 consisted of those in Beijing Municipal City, Tianjin Municipal City, Hebei Province, Shandong Province, north Jiangsu Province, and north Anhui Province; Cluster 4 was made up of those in Zhejiang Province, Shanghai Municipal City, south Anhui Province, south Jiangsu Province, and south Jiangxi Province.

Figure 1 National distribution graph of average incidences of anorectal atresia/stenosis in different provinces, municipal cities and autonomous regions from 2001 to 2005.

Figure 2 Two-dimensional MST-based cluster graph of monitoring sites in China from 2001 to 2005.
north Hunan Province, north Jiangxi Province, Hubei Province, Henan Province, Shanxi Province and Inner Mongolia Autonomous Region; Cluster 5 consisted of those in Ningxia Hui Autonomous Region, Gansu Province and Qinghai Province; and Cluster 6 included those in Shaanxi Province, Sichuan Province, Chongqing Municipal City, Yunnan Province, Guizhou Province, Xinjiang Uygur Autonomous Province and Tibet Autonomous Region (Figure 3).

DISCUSSION
Anorectal atresia/stenosis is one of the most common malformations in the gastrointestinal tract. Due to pathological changes in the anus and rectum, one-third of the perinatal births with anorectal atresia/stenosis suffer from defecation difficulties of varying degrees following surgery. Most of these births need life-long treatment that severely compromises the quality of life and psychological development in particular. This situation is a burden not only to these babies, but also to their entire families and even to society as a whole in China[13-17]. Some researchers[18-22] suggested that mothers’ contact (when they are pregnant) with environmental pollutants could increase their risk of giving birth to babies having congenital malformations. The current research found that the areas with the highest incidences of anorectal atresia/stenosis were concentrated in Eastern China, especially in Liaoning, Zhejiang and Guangdong. With a solid industrial and agricultural base, economic conditions in Eastern China flourish. Most manufacturing plants and industrial factories (including marine-aquatic industries) are located in Eastern China. It is known that these factories are responsible for water pollution and other industrial pollution at a level that is deemed severe. Perhaps mothers in Eastern China have babies with more congenital malformations because of the mothers’ severe exposure to these physical and chemical pollutants when they are pregnant. In addition, the regional differences in awareness and uptake of available health care for pregnant woman, infrastructure of monitoring hospitals and diagnosis at a technical level were also factors likely to explain some of the observed geographical variation in anorectal atresia/stenosis. In Western China, limited at economic and cultural levels, most pregnant woman have weak awareness and uptake of health care, which means they do not actively seek antenatal care, so there is the probability of underreporting of cases, resulting in the lower incidence. As to the health services, in the less developed western regions, the maternal and child healthcare facilities may lack necessary infrastructure, and the technical levels of monitoring staff may be limited, which may also result in the lower detection of congenital malformation.

Cluster analysis is an exploratory data analysis tool for solving classification problems. Assuming the samples as the vectorial points in hyperspace, the object of cluster analysis is to sort the samples into clusters so that the degree of association is strong between members of the same cluster and weak between members of different

Figure 3  Cluster graph of monitoring sites in China from 2001 to 2005.
monitoring sites of the CBDMN into different clusters, the detailed relevant environmental risk factors for anorectal atresia/stenosis in different geographic units can be collected within the same cluster to allow regional monitoring.

The current research took account of the adjacent relationship between different monitoring sites rather than different provinces, autonomous regions or municipal cities, which guaranteed the requirements for geographic divisions for this study. However, if different monitoring sites in the same province were incorporated into different clusters after two-dimensional graph-theoretical clustering, the monitoring work at the provincial level would be subjected to increased difficulties.

ACKNOWLEDGMENTS

We thank monitoring hospitals of the CNBDMN for data collection of anorectal atresia/stenosis. We also thank Mary Meyer and Steven Pan, who polished the paper with meticulous efforts.

REFERENCES

1 International Clearinghouse for Birth Defects Surveillance and Research (ICBDSR). Annual Report 2006 with data for 2004. Roma: The International Centre on Birth Defects, 2006
2 Zhu J. Birth defect and its surveillance. Zhongguo Shiyong Fuke Yu Chunke Zazhi 2002;18: 513-514
3 Dai L, Zhu J, Zhou G, Wang Y, Wu Y, Miao L, Liang J. [Dynamic monitoring of neural tube defects in China during 1996 to 2000] Zhonghua Yu Fang Yi Xue Za Zhi 2002; 36: 402-405
4 Douglas BW. Introduction to Graph Theory. 2nd edition.
How to divide Eastern, Middle and Western China.

Fang KT, Pan EP. Cluster analysis. Beijing: Geographic Publishing House, 1982: 116-119

Page RDM. Graphs and generalized tracks, quantifying Croizat’s panbiogeography. Systematic Zoology 1987; 36: 1-17

Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 1964; 29: 1-27

Tang QY, Feng MG. DPS Data Processing System for Practical Statistics. 1st edition. Beijing: Science Press, 2002: 265-267

Fang SH. Graph theory. Beijing: Science Press, 2004

Xu JM. Graph theory and its application. Hefei: University of Science and Technology of China Press, 2008

Dunn G, Everitt BS, Cunnings C. An introduction to mathematical taxonomy. Cambridge: Cambridge University Press, 1982

Page RDM. Graphs and generalized tracks, quantifying Croizat’s panbiogeography. Systematic Zoology 1987; 36: 1-17

Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 1964; 29: 1-27