# Chinese National Twin Registry as a Resource for Genetic Epidemiologic Studies of Common and Complex Diseases in China

Huiying Yang[1,2], Xiaohui Li[1,2], Weihua Cao[2], Jun Lu[2], Tao Wang[2], Siyan Zhan[2], Yonghua Hu[2], and Liming Li[2,3]

[1]*Division of Medical Genetics, Cedars-Sinai Medical Center, UCLA School of Medicine, USA*
[2]*School of Public Health, Health Science Center, Peking University, China*
[3]*Chinese Center for Disease Control, China*

Twins, due to their unique genetic and environmental relationships, have provided crucial insight in our understanding of genetic contributions to numerous etiologically complex disorders in developed countries. As the leading cause of death and adult disability, cardio- and cerebrovascular diseases are common in China, followed by cancer. Obesity and psychological disorders are increasing. The overall goal of this program is to develop a resource for genetic epidemiologic studies of these and other common and complex diseases in China. Our initial focus is to delineate the genetic and environmental determinants of vascular diseases in general, coronary artery disease (CAD) and stroke in particular. To date, we have over 4500 twin pairs registered and about 700 twin pairs studied for various metabolic traits (e.g., lipids, glucose, insulin, etc.). The long-term plan of this program is to (1) establish a population-based twin registry from several selected regions in China for future studies of specific common complex diseases; (2) conduct detailed phenotyping for clinical and intermediate traits related to cardiovascular diseases; (3) expand studies of twins to twin families by including their parents, siblings, and offspring for genetic linkage and association studies; and (4) follow up twins in the registry longitudinally. The goals of the program are health education and promotion of healthy behavior, early identification of cases to provide timely medical attention, and the evaluation of long-term effects of identified risk factors. We want to develop collaborations with investigators who have expertise in cancer, psychological disorders, and other disease areas.

Family studies of human traits can help answering three questions: first, whether a given phenotypic trait is genetically influenced; second, what is the mode of inheritance? And third, where on the human chromosomes disease susceptibility or trait-influencing genes are located and what specific genetic variants are responsible? Twin data are traditionally used to answer the first question by measuring the relative contribution of genes and the environment to the phenotype under investigation. With the development of the human genome project, twins can also be used to answer the third question by performing linkage and linkage disequilibrium mapping in a large number of extreme concordant and/or discordant dizygotic twin pairs. Compared with the standard sib-pair design, dizygotic twins have the advantage of being the same age and presumably were exposed to a more similar environment than non-twin sib pairs.

Findings of significant heritability in one population may not necessarily be extrapolated to a second population in which there are differences in exposure to relevant environmental factors. Replication of twin studies in different countries provides one way to study such interactions between genes and environment. Unfortunately, population-based twin registries from Asia, especially China, are currently underrepresented.

Therefore, we began to establish a population-based twin registry in China to provide a long-term special population resource for genetic epidemiologic studies of etiologically complex diseases and related phenotypes. The initial focus will be studying stroke- and CAD-related traits.

Based on a conservative estimate (see below), the total number of existing twins available for study is remarkable. In China, especially in rural areas, the population is relatively stable in comparison with Western countries, and residence information is thoroughly monitored by neighborhood committees, primary health care networks and public security bureaus. Given the necessary resources we propose to achieve the initial goal of registering approximately 45,000 twin pairs in five years. Preliminary data from Qingdao City, Shandong Province has demonstrated feasibility and will serve as a model for other areas.

Cardiovascular and cerebrovascular diseases are major public health problems in China. Epidemiological surveys in China have shown that the prevalence of hypertension has increased 25% from 1980 to 1991, and the current prevalence rate is 13.6% in the general population aged 15 years and older (Wu et al., 1993). The incidence of hypertension is 1%, so there are about 3.5 million new cases every year in China. In both rural and urban areas, stroke is the leading cause of death in China. The death rate from stroke was 134.59/100,000 in urban and 110.92/100,000 in rural

*Address for correspondence: Huiying Yang, Division of Medical Genetics, Cedars-Sinai Medical Center, UCLA School of Medicine, USA. Email: Huiying.Yang@cshs.org*

areas in 1996 (National Bureau of Statistics of China, 1997). Every year there are 1.5 million new cases of stroke. Two thirds of these new cases ended in death, while almost one third resulted in disability, which requires enormous emotional, financial and social support. CAD is another cardiovascular disease with increasing incidence in China. The mean increase rate of the CHD death rate was 6.47% from 1990 to 1995 (National Bureau of Statistics of China, 1991–1996). Thus, prevention of cardiovascular and cerebrovascular diseases is a major public health task in China.

To our knowledge, no large-scale genetic epidemiological study has been conducted for stroke or CAD in China. The NIH-funded large scale and multi-ethnic genetic epidemiological studies in the United States will provide some important etiological insight in these disorders, but Chinese in China have a different genetic background, and are exposed to different environmental risk factors. Comparison between Chinese in China and Chinese Americans may help us untangle the complex interaction between gene and environmental factors in the development of CAD and stroke. Important risk factors for Americans may or may not apply to Chinese in China. Therefore, it is extremely important to conduct large-scale, well-designed, genetic epidemiological studies to investigate both genetic and environmental risk factors and their interactions in the Chinese population in China.

In this article, we will summarize the status of this program after its first year with the financial support from the China Medical Board, New York. No scientific results have been generated from this program yet. However, the valuable experience gained during the initial stage will accelerate the establishment of successful twin registries in China. We provide the results of the registration of twins from one study region, Qingdao (Shandong), and the detailed phenotypic investigations of a subset of those twins.

## Design and Methods

The primary goal of this program is to establish a population-based twin registry from several selected regions representing north, south, urban, and rural areas in China. A secondary goal for the next five years is to study genetic contributions to cardiovascular diseases and related intermediate traits, and to test associations of candidate genes with these phenotypes. Therefore, in our study design we have built in several levels of involvement. All study protocols were reviewed and approved by the Ethics Committee for Human Subject Studies in Peking University Health Science Center.

### Establishment of a Population-based Twin Registry from Four Cities/Provinces

We selected four large provinces/cities from both south (Shanghai, Zhejiang) and north China (Beijing, Shandong). Within each of these provinces/cities we intend to cover 10% of its population. Which villages or neighborhoods will be included will depend on the support of local government, availability of primary care network records, and quality of local medical and public health staff. In order to minimize selection bias, we plan to employ all of the following methods to identify twin pairs from the selected communities. In a particular location certain methods may play a more predominant role than others, but public media will not be the only mechanism used in any region.

***Neighborhood/village committees.*** Neighborhoods (villages in rural areas) are the primary community units, which are managed by the neighborhood/village committees. The committee keeps the household information of all the residents living within its neighborhood/village. The committee's function ranges from collecting water/electricity fees to coordinating census survey because they can contact all households in the neighborhood/village. Usually the committee members are composed of retired workers who have lived in the neighborhood/village for most of their lifetime and know most residents personally.

***Local health care networks.*** These networks are the fundamental mechanisms for government to implement public health related campaigns, such as vaccination, family planning, and disease surveillance. In many cases, the health care network works through neighborhood committee and local hospitals.

***Residence registry in the public security bureaus.*** Public security bureaus keep records on all residents in its serving areas, including birth, death, move-in, -out records. These records are required for school registration and employment. Therefore, Public security bureaus are likely to have complete records on residents in the region. Public security bureaus residence database can be used in a similar fashion as the DMV database used in the United States (Cockburn et al., 2001) to search for twins by matching on residence and birth date.

***Public media.*** Public media includes local newspapers, TV and radio stations. In this twin registry program, the public media will be used to educate the public and to solicit volunteers along with other methods mentioned above.

Demographic variables (birth date, gender, ethnicity, occupation) and contact information (address and phone number) for twin pairs were obtained from these sources. Only twins with both living co-twins who live in the twin registry coverage areas were included in the registry.

### Detailed Phenotype Studies in a Subset of Twins Aged 25 Years or Older

Selected twin pairs were interviewed in person by community doctors and public health workers. All twin pairs where both twins signed the informed consent form were enrolled for participation. Specifically they completed the following protocols:

***Questionnaire.*** All recruited twins were interviewed by community doctors and staff. Questionnaire information included:

1. Questions designed to predict whether the twins are identical (monozygotic) or fraternal (dizygotic) and whether they have shared the same environment as they grew up.
2. Demographic information — birth date, gender, education, occupation, marital status.
3. Life style — smoking, drinking, physical activity.

4. Medical history — regarding common chronic diseases, age of onset, treatments.
5. Family history — including parents, siblings, and children: whether they are still living and their residence and presence of common chronic diseases and ages of onset. For the deceased family members, age at death and cause of death were obtained.

***Physical examination and collection of fasting blood sample.*** We measured weight, height, blood pressure, waist/hip ratio, and collected fasting blood samples.

***Biochemical measures.*** Fasting lipids (TC, TG, HDL, LDL), lipoproteins, glucose, insulin, homocysteine were analyzed. All laboratory analyses were performed in Peking University Health Science Centre with WHO approved protocols. All phenotypic results were given to participants and if there were any abnormal findings, clinical advice was provided.

After less than a year of data collection (the official starting date of this project was July 1, 2001), we have generated preliminary data for the initial registry and detailed phenotype studies of a long-term ambitious twin program.

## Results

### Distribution of Estimated Twin Pairs in the Selected Regions

Table 1 lists the cities/provinces we have selected and the estimated numbers of living twin pairs and targeted number of twins based on 10% population coverage. These estimated numbers were based on the twin birth rate, population size and age distribution, and age-specific mortality as reported in the 1990 National Census in China (National Population Census Office of China, 1993).

### Response Rate in Qingdao City, Shandong Province

Qingdao, one of cities in Shandong province, is the first location in which we conducted systematic enrolment based on the study design as discussed above. There are a total of 12 districts in Qingdao city. We have started to ascertain twin pairs in 10 districts (Table 2). The first three districts are considered urban areas with 6% rural population, and the other seven districts are rural areas with 86% rural population. The estimated number of twin pairs in these districts is 16,954 based on its population size. Response rates of the registry by districts range from 10.5% to 81.3% by the end of 2001. Table 2 lists the response rate in each district of Qingdao city. Of 4374 registered twin pairs in Qingdao city, 70% were registered through local health care networks and neighborhood committees. The remaining 30% of twin pairs heard about our study through the public media.

**Table 1**

Geographic Locations and Estimated Number of Twins

| City/ Province | Region | Population Size (million) | Estimated No. of Living Twin Pairs | Targeted No. of Twin Pairs in Registry |
|---|---|---|---|---|
| Beijing | North | 10.8 | 32,832 | 3283 |
| Shandong | North | 84.4 | 256,576 | 25,658 |
| Shanghai | South | 13.3 | 40,432 | 4043 |
| Zhejiang | South | 41.4 | 125,856 | 12,586 |
| Total | | | | 45,570 |

**Table 2**

The Response Rate by District in Qingdao City

| Name of district | Estimated number of twin pairs | Registered number of twin pairs | Response rate (%) |
|---|---|---|---|
| Shinan | 1194 | 172 | 14.4 |
| Shibei | 1329 | 179 | 13.5 |
| Licang | 794 | 194 | 24.5 |
| Laoshan | 544 | 442 | 81.3 |
| Huangdao | 544 | 202 | 37.2 |
| Chengyang | 1214 | 682 | 56.1 |
| Jimo | 3024 | 1105 | 36.5 |
| Jiaozhou | 2148 | 418 | 19.5 |
| Jiaonan | 2376 | 581 | 24.4 |
| Pingdu | 3787 | 399 | 10.5 |
| Total | 16,954 | 4374 | 25.8 |

### Distribution of Registered Twin Pairs and Participation in Detailed Phenotypic Studies in Qingdao city, Shandong Province

The distribution of registered twins by age and gender is listed in Table 3.

Out of 4374 twin pairs, 1248 are aged 25 or older. Among these 1248 twin pairs, 510 pairs were enrolled in the detailed phenotypic study by the end of 2001. The proportion enrolled for each age group is given in Table 4.

The main reason for non-enrolment was that the field work was conducted on limited days and hours during the day (in the morning to obtain fasting blood) and many twins had difficulty fitting these fixed times into their work schedule. In addition, if only one twin could make it and the other could not, these twin pairs were not included in the registry nor in the detailed study. In general, twins were enthusiastic about this project; they appreciated the special attention paid to this special population and expressed the view that it was beneficial to receive a free physical examination and

**Table 3**

The Age and Sex Distribution of Twins in the Chinese Twin Program up to the End of 2001

| Region | Birth year | Twins (Pairs) | | | Total |
|---|---|---|---|---|---|
| | | Male | Female | Mixed | |
| North: | < 1929 | 11 | 5 | 2 | 18 |
| Qingdao, | 1930–39 | 13 | 7 | 6 | 26 |
| Shandong | 1940–49 | 33 | 12 | 10 | 55 |
| Province | 1950–59 | 119 | 40 | 40 | 199 |
| | 1960–69 | 253 | 136 | 112 | 501 |
| | 1970–79 | 339 | 279 | 157 | 775 |
| | 1980–89 | 438 | 435 | 245 | 1118 |
| | 1990–99 | 662 | 597 | 344 | 1603 |
| | 2000– | 34 | 28 | 17 | 79 |
| | Total | 1902 | 1539 | 933 | 4374 |

**Table 4**

Proportion of Registered Twin Pairs Participated in Detailed Study by the End of 2001

| Birth year | Registered* | Enrolled in detailed study (%) |
|---|---|---|
| < 1929 | 18 | 3 (16.7) |
| 1930–39 | 26 | 8 (30.8) |
| 1940–49 | 54 | 27 (50.0) |
| 1950–59 | 197 | 121 (61.4) |
| 1960–69 | 492 | 182 (37.0) |
| 1970–75 | 461 | 169 (36.7) |
| Total | 1248 | 510 (40.9) |

Note: * Registered number did not include 30 twin pairs recruited during the pilot stage.

**Table 5**

The Distribution of Zygosity by Age and Sex

| Birth Year | MZ (pairs) | | | DZ (pairs) | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | Male | Female | Total | Male | Female | Mixed | Total | |
| < 1929 | 2 | 0 | 2 | 1 | 0 | 0 | 1 | 3 |
| 1930–39 | 5 | 1 | 6 | 1 | 1 | 3 | 5 | 11 |
| 1940–49 | 16 | 7 | 23 | 1 | 0 | 7 | 8 | 31 |
| 1950–59 | 43 | 54 | 97 | 10 | 7 | 22 | 39 | 136 |
| 1960–69 | 79 | 104 | 183 | 9 | 12 | 47 | 68 | 251 |
| 1970–79 | 77 | 90 | 167 | 17 | 10 | 40 | 67 | 234 |
| 1980–89 | 29 | 25 | 54 | 8 | 7 | 7 | 22 | 76 |
| Total | 251 | 281 | 532 | 47 | 37 | 126 | 210 | 742 |

metabolic blood tests. Participants understood that this was voluntary and some twins did refuse to participate.

### Zygosity of Enrolled Twins

By the end of 2001, besides 510 twin pairs enrolled from Qingdao, there were an additional 232 twin pairs recruited during the pilot stage for detailed phenotypic studies (202 from Lishui, Zhejiang province and 30 from Shandoing province). Therefore, we have collected a total of 742 twin pairs with detailed phenotypic information and blood samples. To determine zygosity, the ABO blood type was used for an initial screen, then DNA tests with higher specificity (99.99%) were performed. The distribution of zygosity by age and gender is listed in Table 5. The zygosity distribution in this data set is consistent with that reported for the region (Gan & Zheng, 2001).

## Discussion

During the first year of a 5-year project, we have registered 4576 living twin pairs (including 4374 from Qingdao and 202 selected samples from Lishui) and collected detailed phenotype data and DNA samples on 742 twin pairs. Our goal is to register 45,570 twin pairs in 4 provinces/cities of China. To achieve this goal we may have to expand coverage areas from 10% to 13–17% of the selected regions assuming 20–40% non-response rate. Our first year consisted of protocol developments, initial training of staff, and recruitment of twins. Now with the experience learned from the Qingdao site, and with trained staff, our registration rate will accelerate in the next four years. The response rate ranges from 10.5% to 81.3% across 10 districts in Qingdao city. These rates represent the work in progress after one year of ascertainment effort. We expect that the final response rate would be close to the higher value of this range by the end of this project.

The goal of studying adult twins on cardiovascular disease associated metabolic traits is to estimate the genetic heritability of these traits in this population in order to determine phenotypes useful for further gene identification studies. The logic here is that if a trait is mainly influenced by environmental factors, the effort to identify susceptibility genes may be too great to be successful. Therefore, we will initially focus on those traits with significant genetic contribution for family studies as stated in the future objectives of this program.

The following two aims are designed to be part of this program and will be pursued at a later stage.

1. Expand twin studies to twin families by including their parents, siblings, and offspring for genetic linkage and association studies to identify disease susceptibility genes to risk conditions (e.g., elevated blood pressure, adverse lipid profile, and abnormal glucose level) for CAD and/or stroke with the candidate gene study approach.
2. Follow up twins in the registry longitudinally by various schemes based on age and risk factors (a) to provide health education and promote healthy behavior, (b) to identify early cases and provide timely medical attention, and (c) to evaluate long term effects of identified risk factors for CAD and stroke.

During this registration stage, we collected information about families and their medical histories to prepare for future disease specific genetic epidemiological studies. This initial success of the twin registry and the collection of phenotypic data and DNA samples provide us with an optimistic projection for continued success.

We realize that the phenotypes studied at this stage are not comprehensive for studying either CAD or stroke, but this is the beginning of an expandable genetic epidemiology program. We can expand in CAD or stroke related areas (such as studying Carotid intima-medial wall thickness, body fat measurement) and/or to other diseases, such as cancers, obesity, osteoporosis, autoimmune diseases, and psychiatric disorders.

Any potential collaboration proposals are welcome and will be evaluated by investigators in this program on their scientific merit, relevance to the Chinese population, and availability of resources.

## Acknowledgment

The authors want to express our heartfelt thanks to the China Medical Board, New York for its financial support (CMB01-746), to local public health and medical workers, and especially to those study participants without whom this study would not be possible. Special thanks to Drs. Zengchang Pang and Tao Bo from Qingdao's Center for

Disease Control for their dedication and commitment to this project, and to Drs. Myles Cockburn and Jerome I. Rotter for their helpful discussion and enthusiastic support of this program.

## References

Cockburn, M. G., Hamilton, A. S., Zadnick, J., Cozen, W., & Mack, T. M. (2001). Development and representativeness of a large population-based cohort of native Californian twins. *Twin Research, 4,* 242–250.

Gan, J., & Zheng, J. (2001). The influence of geological factor on the birth rates of twins in China. *Journal of Central China Normal University: Nature Science Version, 35*(3), 338–341.

National Bureau of Statistics of China. (1991–1997). *China Statistical Yearbook (1991–1997).* Beijing: National Bureau of Statistics of China Press

National Population Census Office of China. (1993). *The 1990 National Census Databook*. Beijing: Chinese Statistic Press.

Wu, X. G., He, J. S., Fang, W. C., Dang, S. F., Fang, D. Z., Liu, S. Z. et al., (1993). *The 1991 National Hypertension Survey Databook* [in Chinese]. Beijing: Peking Union Medical College Press.