Determination of Zygosity by Questionnaire and Physical Features Comparison in Chinese Adult Twins

Wenjing Gao,[1] Liming Li,[1] Weihua Cao,[1] Siyan Zhan,[1] Jun Lv,[1] Ying Qin,[1] Zengchang Pang,[2] Shaojie Wang,[2] Weijian Chen,[3] Rongfu Chen,[3] and Yonghua Hu[1]

[1] *Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China*
[2] *Center for Disease Control and Prevention, Qingdao, China*
[3] *Center for Disease Control and Prevention, Lishui, China*

This study reports on the determination of zygosity in Chinese adult twins by simple questionnaire and physical features comparison. The subjects were 511 twin pairs from two cities and their town areas, consisting of 371 monozygotic (MZ) and 140 same-sex dizygotic (DZ) pairs, identified by ABO blood group and multiplex polymerase chain reaction of several polymorphic short tandem repeat markers. The twins themselves responded to 8 questionnaire items, 4 items on twin similarity, and 4 items on the frequency of mistaking one twin for another by parents, relatives, teachers and strangers when they were 6 to 13 years old. Research assistants responded to 20 items regarding twins' physical features at the moment of interview. A parsimonious model established using stepwise logistic regression analysis of the 28 items showed that the total accuracy of zygosity diagnosis was 90.1%. The accuracy was 89.2% when using only the items dealing with the confusion of twins and 85.4% using only similarity. In the questionnaire, 'facial appearance', 'mistaken by teachers' and 'mistaken by strangers' had stronger discriminating power between MZ and DZ twins. Two physical features — 'eyelid' and 'middigital hair' — were informative to some extent. There was no statistically significant sex and area difference in the validity of such questionnaire and physical features comparison-based classification. In conclusion, questionnaire-based zygosity assessment in this Chinese adult twin sample could still be regarded as a valid and valuable classification method. Physical features comparison, however, could only provide limited information for zygosity determination.

Classical twin design relies upon the comparison of concordance in monozygotic (MZ) and dizygotic (DZ) twins. Whether zygosity classification is correct or not may have a large effect on the value of twin study. In the excitement about the advances in molecular genetics, from the late 1980s, DNA analysis has been the most accurate method for zygosity diagnosis, which allows a level of false classification to be close to zero and is considered to be the current 'gold standard'. However, the expense and the time-consuming nature of genetic testing make it infeasible in large-scale epidemiological studies.

Assessing the extent to which co-twins look like each other using questionnaires is another method for zygosity classification. Many twin studies have shown that questionnaire-based zygosity diagnosis can achieve accuracy of around 95% (Rietveld et al., 2000). It has been widely used for its ease and low cost.

In the last decade, the validity of questionnaires for young twins has attracted more and more researchers (Chen et al., 1999; Forget-Dubois et al., 2003; Jackson et al., 2001; Ooki & Asaka, 2004; Price et al., 2000; Rietveld et al., 2000; Spitz et al., 1996), while for adult twins, with a rapidly increasing need for an appropriate zygosity diagnosis method, it has been gradually ignored. The growth in need could be attributed to three reasons as follows: first, adult twins are excellent subjects for the genetic epidemiological study of complex diseases that are increasingly becoming the focus of attention in disease control and prevention. Second, with progress in molecular genetics, it is time for questionnaires, designed for adult twins many years ago, to be validated by the more accurate 'gold standard'. Third, the questionnaires for zygosity determination are mainly based on twins' physical similarity. However, physical features, such as eye colour and hair colour, are different in different races. To our knowledge, especially in Chinese adult twins, the validity of specific physical features comparison-based zygosity determination has not been reported. The

Received 2 November, 2005; accepted 14 December, 2005.

Address for correspondence: Wenjing Gao, Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, 38 Xueyuan Road, Haidian District, Beijing 100083, China. E-mail: gaowenjing1979@gmail.com

purpose of the present study therefore is to use the current DNA zygosity classification in the Chinese National Twin Registry to compute the validity of observations of twins' physical features combined with questionnaires in zygosity determination.

Materials and Methods

Subjects

The Chinese national population-based twin registry system (Yang et al., 2002) was established in 2001 and in the last 5 years, August 2001 to March 2005, more than 6000 twin pairs of all age and sex have been enrolled. Based on the registry, a twin cohort of 1008 pairs were established in two areas — Qingdao, Shandong province and Lishui, Zhejiang province — the database of which includes zygosity questions, other detailed information and a 10 ml fasting blood sample after obtaining their written informed consent. The twin pairs were recruited through local general practitioners and centers for disease control and prevention. During the follow-up of these twins 2 to 3 years later, a total of 511 same-sex twin pairs (58 MZ male, 92 MZ female, 26 DZ male, and 42 DZ female pairs from Qingdao and 128 MZ male, 93 MZ female, 44 DZ male, and 28 DZ female pairs from Lishui) received a physical features observation by trained research assistants. The twins' mean age was 40.9 ± 9.9 years (range = 21.5–83.3 years). Zygosity in 218 pairs from Qingdao was classified by ABO blood group and the multiplex polymerase chain reaction (PCR) of nine polymorphic short tandem repeat (STR) markers (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820) and Amelogenin locus using PE AmpFISTR Profiler Plus™ Kit. Two hundred and ninety-three twin pairs from Lishui were classified by ABO blood group and four highly polymorphic STR markers (D16S539, D7S820, D13S317 and D5S818) using Promega GammaSTR® Multiplex (Flurescein) Kit. Using these two marker sets the probability of monozygosity determined by identity of all markers is estimated to be at least .999 (Lv et al., 2003) and .996 (Chen et al., 2004) respectively.

Zygosity Diagnosis

We constructed two questionnaires for the twins themselves and trained research assistants respectively. The first questionnaire for twins self-reporting included 8 items regarding physical similarity and the confusion of twins. First, twins were asked whether they were similar in terms of (a1) facial appearance, (a2) height, (a3) weight, (a4) hair texture when they were between 6 and 13 years old, with three possible answers to (a1): 'they were as alike as two peas in a pod', 'just as alike as two ordinary siblings' and 'hard to say'. Possible answers to items a2 to a4 were 'only a slight difference', 'clear difference' and 'hard to say'. In the following 4 items regarding confusion, twins were asked whether they were ever mistaken for one another by (b1) parents, (b2) other relatives and friends, (b3) teachers and (b4) strangers, when they were between 6 and 13 years old. The three possible answers to these 4 items were 'yes', 'no' and 'hard to say'.

The second questionnaire consisted of 20 physical feature items for research assistants, who were blind to twins' DNA diagnosis of zygosity. To minimize the subjectiveness of research assistants' answers, the questionnaire did not ask the extent to which the twins looked like each other, but rather research assistants were asked to describe twins' physical features according to a definition (see Appendix A), for which they had been trained prior to interviewing the twins. If a feature on the left side was inconsistent with that on the right, research assistants would select the dominant trait as the answer. A simian crease on either the left or right hand was defined that he or she had simian crease. Finally, answers on the 20 items in each pair were compared to determine whether they were different.

Results

For each of the 28 items in the two questionnaires, the frequency of responses or the results of comparison were compared with the DNA diagnosis of zygosity (Table 1). To improve the accuracy of DZ, a 2-point scale was used for coding (1 = 'they were as alike as two peas in a pod' for item a1, 'only a slight difference' for a2 to a4, 'yes' for b1 to b4; 2 = 'just as alike as two ordinary siblings' for a1, 'clear difference' for a2 to a4, 'no' for b1 to b4 and 'hard to say' for all items regarding the similarity and confusion of twins. For items c1 to c20, 1 = same answer for Twin 1 and Twin 2; 2 = different answers for Twin 1 and Twin 2. In 18 items (a1–a4, b1–b4 and c1, 4, 7, 8, 10, 11, 15, 16, 17, 19), we found statistically significant agreements between their determination and DNA classification at the .05 level. But the majority of these items in our questionnaires, especially regarding features, had little discriminating power between MZ and DZ as very few twins had any difference in these answers. Only 'facial appearance' of the similarity items and 'mistaken by other relatives and friends', 'teachers', 'strangers' of the confusion items had a total accuracy of more than 80.0%. The DZ accuracy of all these 28 items was not as high as the accuracy of MZ twins.

To construct a parsimonious model predicting twin zygosity among the 28 items, we employed stepwise logistic regression analysis with a significance level of .10 for entry or staying in the model (SAS Institute, 1993). For the questionnaire regarding similarity, items a1, 2, 4 ('facial appearance', 'height', 'hair texture') were selected in the final model. For the questionnaire regarding confusion, items b3 and b4 ('mistaken by teachers' and 'strangers') were selected, whereas the final model for the items regarding features selected items c1, 7, 8, 10, 11, 15, 17 ('hair', 'ear lobes', 'earwax', 'nostril shape', 'tongue rolling', 'Hitchhiker's thumb' and 'middigital hair'). The regression coefficients and prediction accuracy indexes are listed in Table 2. Model IV was for all the items included in the questionnaires regarding (a) similarity, (b) confusion and (c) features, that is, a + b + c. Model V included a + b + c

https://doi.org/10.1375/twin.9.2.266 Published online by Cambridge University Press

Table 1

Individual Item Accuracy in Determining Zygosity Classification

	Number of*		Accuracy		Total	Kappa†	z‡	p value
	MZ	DZ	MZ (%)	DZ (%)				
a. Similarity								
a1. Facial appearance	370	140	90.3%	70.7%	84.9%	.617	13.931	< .001
a2. Height	370	140	95.4%	37.9%	79.6%	.394	9.742	< .001
a3. Weight	370	140	92.7%	37.9%	77.6%	.353	8.469	< .001
a4. Hair texture	371	139	98.4%	28.8%	79.4%	.343	9.534	< .001
b. Confusion								
b1. Parents	369	140	13.0%	97.1%	36.1%	.060	3.377	.001
b2. Other relatives and friends	370	140	85.1%	75.0%	82.4%	.576	13.061	< .001
b3. Teachers	369	140	79.9%	80.7%	80.2%	.549	12.676	< .001
b4. Strangers	371	140	97.0%	68.6%	89.2%	.708	16.256	< .001
c. Features comparison								
c1. Hair	370	140	98.6%	7.9%	73.7%	.090	3.761	< .001
c2. Widow's peak	371	140	81.1%	26.4%	66.1%	.082	1.873	.061
c3. Number of hair whorls	370	139	90.8%	7.9%	68.2%	–.016	–0.452	.652
c4. Mongoloid fold	371	140	97.0%	7.9%	72.6%	.066	2.430	.015
c5. Double eyelid	371	140	85.2%	20.0%	67.3%	.059	1.415	.157
c6. Darwin's tubercle	371	140	96.0%	6.4%	71.4%	.032	1.137	.256
c7. Ear lobes	371	140	89.5%	25.7%	72.0%	.178	4.331	< .001
c8. Earwax	371	140	97.0%	12.9%	74.0%	.131	4.311	< .001
c9. Aquiline nose	371	140	99.5%	0.7%	72.4%	.003	0.231	.817
c10. Nostril shape	371	140	92.7%	20.7%	73.0%	.165	4.337	< .001
c11. Tongue rolling	371	140	82.7%	29.3%	68.1%	.131	3.003	.003
c12. Arm-folding	369	140	52.8%	51.4%	52.5%	.035	0.862	.389
c13. Handedness	371	140	87.3%	12.9%	66.9%	.002	0.057	.954
c14. Hand-clasping	370	138	54.6%	49.3%	53.1%	.032	0.778	.437
c15. Hitchhiker's thumb	367	138	90.2%	19.6%	70.9%	.117	2.957	.003
c16. Forefinger and ring finger, which longer?	363	132	93.1%	12.9%	71.7%	.077	2.116	.034
c17. Middigital hair	371	140	97.0%	11.4%	73.6%	.113	3.814	< .001
c18. Bent pinky	370	140	98.4%	2.9%	72.2%	.017	0.898	.369
c19. Simian crease	369	139	95.7%	9.4%	72.0%	.067	2.173	.030
c20. Daltonism	366	138	100.0%	0.0%	72.6%	—	—	—

Note: * The MZ and DZ twin pairs were determined by DNA. Some were deleted because of the target variable missing.
† Kappa measures the agreement between the item determination and DNA classification.
‡ The z statistic tests for the null hypothesis that there is no agreement.

with twins' sex, geographical location, and the time twins had spent together with each other, but the variables for sex and geographical location were not selected in it. If the significance level of entry into or staying in the model was changed from .10 to .05, the results remained the same except that a4, c8, c10 ('hair texture', 'earwax' and 'nostril shape') were not retained in the final Model IV and 'time', a4, c8, c10 were not in Model V. The total predictive accuracy of the models was better for both the questionnaires regarding the similarity and confusion of twins (85.4% and 89.2%, respectively), with a sensitivity of about 70% and a specificity of greater than or equal to 90.0% in both models. However, the total predictive accuracy of the model for questionnaire regarding features comparison was less satisfactory (75.9%), with a sensitivity of only 26.3%.

In general, a1, b3, b4, c5, c17 ('facial appearance', 'mistaken by teachers' and 'strangers', 'difference in eyelid' and 'middigital hair') together had stronger discriminating power between MZ and DZ in this Chinese adult twin sample. With 'hair texture', 'earwax' and 'nostril shape', a total accuracy of 90.1% was achieved (Model IV).

Discussion

Several investigators have confirmed the reliability of the questionnaire method (Chen et al., 1999; Peeters et al., 1998). In our study we used information given by twins themselves and research assistants. The eight items in the first questionnaire were reliable for self-reporting (Chen et al., 1999). A major difference between feature observation by research assistants and

https://doi.org/10.1375/twin.9.2.266 Published online by Cambridge University Press

Table 2

Accuracy of the Parsimonious Logistic Models of Two Questionnaires in Predicting DNA-Determined Zygosity

No.	Questionnaire	Number of[**]		Logistic model[§]	Prediction accuracy			
		MZ	DZ	$\text{Log}[p/(1-p)]$	Cut-off point	Correct	Sensitivity	Specificity
I	a. Similarity	369	139	$-7.2252+2.6076I_{a1}+0.9336I_{a2}+1.4782I_{a4}$	.24–.26	85.4%	73.4%	90.0%
II	b. Confusion	367	140	$-6.9351+1.0744I_{b3}+3.4764I_{b4}$	.22–.88	89.2%	68.6%	97.0%
III	c. Features	356	133	$-1.8645+1.3182I_{c1}+1.1041I_{c7}+1.6625I_{c8}+1.1277I_{c10}+0.6029I_{c11}+0.8813I_{c15}+1.6771I_{c17}$	.42[††]	76.9%	45.9%	88.5%
IV	a + b + c	351	132	$-9.2514+1.3493I_{a1}+1.1090I_{a4}+0.9243I_{b3}+2.6902I_{b4}+0.7747I_{c5}+1.3137I_{c8}+0.9248I_{c10}+2.0067I_{c17}$	.64–.72	90.1%	69.7%	97.7%
V	a + b + c with sex, geographical location and time together	347	130	$-10.5883+0.0444t+1.3013I_{a1}+1.3971I_{a4}+1.0223I_{b3}+2.6462I_{b4}+0.8185I_{c5}+1.3364I_{c8}+0.8666I_{c10}+1.8934I_{c17}$	.74[‡‡]	89.7%	68.5%	97.7%

Note: § p = the probability of being DZ; I_n = item n of the questionnaire; t = the time twins had spent together.
** Several MZ and DZ twin pairs were deleted from the regression as one or more variables were missing.
†† For the cut-off point .42 to .56, the corresponding correct waves have narrow range 76.3% to 76.9%.
‡‡ For the cut-off point .32 to .84, the corresponding correct waves have narrow range 89.1% to 89.7%.

a self-reporting questionnaire for twins was that the observations were based on the features of twins at the time of interview. Disadvantages are the possible influences of observer bias. To minimize the bias and ensure twin–co-twin comparison, the research assistants had been strictly trained with the standard definition of each physical feature and were required to observe both twins at the same time.

In terms of predicting zygosity, it was hoped that as many features as possible were used to improve the accuracy of zygosity determination on the basis of the questionnaire-based method. However, as the results showed, many individual items were useless. As no person in our sample suffered from daltonism, the item about daltonism had no discriminating power. Individual items that had good predictive accuracy in this study were mainly items related to facial appearance similarity and teacher and stranger twin confusion.

Several parsimonious models of more than one item for Parts A, B and C were derived from logistic regression analysis that led to an improvement in predictive accuracy. The predictive accuracy of these models was different for each of the part. The total accuracy of A, B and C was 90.1% with 85.4% for Part A and 89.2% for Part B. Physical feature comparison can only provide limited information for the diagnosis of zygosity. Such a result was comparable to a previous study of Chinese young twin sample (Chen et al., 1999), which also found several physical characters — weight, height, hair texture, shape of ear lobes, hair whorl, thumb curvature, palmar creases, and so forth — did not have sufficient concordance rates with the DNA diagnosis. This may be attributed to the considerably less genetic information in individual features than in the impression as a whole of ‘as alike as two peas in a pod’ or ‘confusion by teachers and strangers’.

Our results indicate that neither individual items nor a parsimonious model of items on features could predict zygosity with sufficient accuracy in this Chinese adult twin sample, which suggests that, when assessing zygosity in Chinese adult twins, feature observation by researchers may be not necessary, although it seems to make the questionnaire more objective.

Twins’ sex and geographical location was not retained in the last logistic model, which shows that sex and geographical location may have little effect on zygosity determination with questionnaire and physical feature observation. The regression coefficient .0444 for time twins spent together meant that with the passage of time, the probability of ‘being determined as DZ’ increased. The accumulation of environmental difference between twins may make them less and less alike. From the regression coefficients of other items, we could also see that the possible effect of such accumulation on twin zygosity diagnosis was not very large.

In conclusion, questionnaire-based zygosity assessment in Chinese adult twins can still be regarded as a valid and valuable classification method. The accuracy of physical features comparison, however, was less satisfactory and needs further independent samples to validate it.

Acknowledgments

This study was funded by grants from National Natural Science Foundation of China (30070672) and China Medical Board (01-746). We thank Canqing Yu, Xun Tang, Na Li, YanHong Wang, Bo Wang, YanPing Zhu, Yan Zhang and Aiqun Huang in data collecting and the help of Prof Jun Ma in study planning and Pingping He in data analysis. We would also gratefully acknowledge the help of Forensic Medical Examination & Identification Center of Beijing Public Security Bureau in DNA zygostiy determination.

References

Chen, T., Ji, C., Pang, Z., Wang, S., Hu, Y., & Qin, Y. (2004). Heritability of serum lipids and lipoproteins and its related factors in twins aged 5 to 19 years of China. *Chinese Journal of Preventive Medicine, 38,* 237–239.

Chen, W. J., Chang, H. W., Wu, M. Z., Lin, C. C. H., Chang, C., Chiu, Y. N., & Soong, W. T. (1999). Diagnosis of zygosity by questionnaire and polymarker polymerase chain reaction in young twins. *Behavior Genetics, 29,* 115–123.

Forget-Dubois, N., Pérusse, D., Turecki, G., Girard, A., Billette, J., Rouleau, G., Boivin, M., Malo, J., & Tremblay, R. E. (2003). Diagnosing zygosity in infant twins: physical similarity, genotyping, and chorionicity. *Twin Research, 6,* 479–485.

Jackson, R. W., Snieder, H., Davis, H., & Treiber, F. A. (2001). Determination of twin zygosity: A comparison of DNA with various questionnaire indices. *Twin Research, 4,* 12–18.

Lv, J., Zhan, S., Qin, Y., Li, Q., Guo, X., Cao, W., Pang, Z., Bo, T., Hu, Y., Li, L., Liu, Y., Huo, Z., & Liang, G. (2003). Determination of twin zygosity in epidemiological studies. *Journal of Peking University (Health Sciences), 35,* 212–214.

Ooki, S., & Asaka, A. (2004). Zygosity diagnosis in young twins by questionnaire for twins' mothers and twins' self-reports. *Twin Research, 7,* 5–12.

Peeters, H., Gestel, S. V., Vlietinck, R., Derom, C., & Derom, R. (1998). Validation of a telephone zygosity questionnaire in twins of known zygosity. *Behavior Genetics, 28,* 159–163.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3,* 129–133.

Rietveld, M. J. H., Valk, J. C. V. D., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D.I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research, 3,* 134–141.

SAS Institute, Inc. (1993). *SAS user's guide. Basics and statistics* (Version 6.12). Cary, NC: SAS Institute.

Spitz, E., Moutier, R., Reed, T., Busnel, M. C., Marchaland, C., Roubertoux, P. L., & Carlier, M. (1996). Comparative diagnoses of twin zygosity by SSLP variant analysis, questionnaire, and dermatoglyphic analysis. *Behavior Genetics, 26,* 55–63.

Yang, H., Li, X., Cao, W., Lu, J., Wang, T., Zhan, S., Hu, Y., & Li, L. (2002). Chinese National Twin Registry as a resource for genetic epidemiologic studies of common and complex diseases in China. *Twin Research, 5,* 347–351.

Appendix A

List of physical features

c1.	Hair	__straight hair	__curly hair	__ hard to say
c2.	Widow's peak§§	__ present	__ absent	__ hard to say
c3.	Number of hair whorls	__ 1	__ 2	__ 3 or more
c4.	Mongoloid fold***	__ present	__ absent	__ hard to say
c5.	Double eyelid	__ yes	__ no	__ hard to say
c6.	Darwin's tubercle†††	__ present	__ absent	__ hard to say
c7.	Ear lobes‡‡‡	__ attached	__ detached	__ hard to say
c8.	Earwax	__ dry	__ sticky	__ hard to say
c9.	Aquiline nose	__ yes	__ no	__ hard to say
c10.	Nostril shape	__ broad	__ narrow	__ hard to say
c11.	Tongue rolling§§§	__ yes	__ no	__ hard to say
c12.	Arm-folding	__ left-sided preference	__ right-sided preference	__ hard to say
c13.	Handedness	__ left-sided preference	__ right-sided preference	__ hard to say
c14.	Hand-clasping	__ left-sided preference	__ right-sided preference	__ hard to say

https://doi.org/10.1375/twin.9.2.266 Published online by Cambridge University Press

c15.	Hitchhiker's thumb****	__ present	__ absent	__ hard to say
c16.	Forefinger and ring finger, which longer?	__ forefinger	__ equal	__ ring finger
c17.	Middigital hair††††	__ present	__ absent	__ hard to say
c18.	Bent pinky‡‡‡‡	__ present	__ absent	__ hard to say
c19.	Simian crease§§§§	__ present	__ absent	__ hard to say
c20.	Daltonism	__ yes	__ no	__ hard to say

Note: §§ Widow's peak: This is a descending V-shaped point in the middle of the hairline (above the forehead). Caused by dominant allele.

*** Mongoloid fold (epicanthic fold): This is a skin fold of the upper eyelid (from the nose to the inner side of the eyebrow) covering the inner corner (medial canthus) of the eye. Caused by dominant allele.

††† Darwin's tubercle: This is a congenital ear condition which often presents as a thickening on the helix at the junction of the upper and middle thirds. Caused by dominant allele.

‡‡‡ Ear lobes may be either adherent or free and pendulous. Homozygous recessives have attached ear lobes; heterozygous or homozygous dominant individuals have detached (free) ear lobes.

§§§ Tongue rolling: Persons with a dominant allele in heterozygous or homozygous condition can roll their tongues into a tube-like shape; homozygous recessives are nonrollers and can never learn to roll their tongues.

**** Hitchhiker's thumb: Homozygous recessives can bend the distal joint of the thumb backward to a nearly 90° angle; heterozygous or homozygous dominant condition yields thumbs that cannot bend backward more than approximately 30°.

†††† Middigital hair: People lacking hair in the middle segments of the fingers are homozygous recessive. The presence of hair on one or more middle segments of the fingers may be governed by a series of alleles each of which is dominant to the recessive.

‡‡‡‡ Bent pinky: Dominant allele causes the distal segment of the fifth finger to bend distinctly inward toward the ring finger.

§§§§ Simian crease: A simian crease is a single palmar crease as compared to two creases in a normal palm.