Optimization of Zygosity Determination by Questionnaire and DNA Genotyping in Chinese Adolescent Twins

Jie Chen,[1] Xinying Li,[1] Zhiyan Chen,[1] Xiaodong Yang,[1] Jie Zhang,[1] Qing Duan[1] and Xiaojia Ge[1,2]

[1] *Key Lab of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China*

[2] *Institute of Child Development, University of Minnesota Twin Cities, Minneapolis 55455, United States of America*

The main aim of this study was to develop and optimize a questionnaire-based zygosity determination method in Chinese adolescent twins. Participants were 471 pairs of same-sex twins (345 monozygotic, 126 dizygotic) with a mean age of 14.56 years (SD = 2.62). A second sample was recruited for cross-validation, including 382 pairs of same-sex twins (261 monozygotic, 121 dizygotic) with a mean age of 12.53 years (SD = 2.22). The questionnaire consisted of 12 questions dealing with co-twin similarity or frequency of confusion. Two means were put forward to improve the predictive accuracy of the questionnaire — adding parent-reports to the analysis, and using a 2-point rather than 3-point response format. DNA genotyping was performed on nine short tandem repeat loci, with an estimated zygosity classification accuracy very close to 100%. The validity of all questionnaires was assessed by being compared to the results of DNA analysis. Results of stepwise logistic regression analysis showed that the predictive accuracy of the 3-point self-reported questionnaire is 83.8%. Using parent-reports and 2-point scale led to 3.9% and 4.6% increase in predictive accuracy, respectively. When using the parent-reports and children's self-reports jointly, the predictive accuracy was enhanced to 90.6%. For the cross-validation, the equations and cut-offs derived from the first sample led to an acceptable accuracy (91.3%) in the second sample. In conclusion, the method we developed can be used in future studies among Chinese adolescent twins. Multiple-rater and 2-point response format were suggested for all twin studies for enhancing the predictive accuracy of questionnaires.

Keywords: zygosity determination in Chinese adolescent twins

Classic twin studies estimate the relative contribution of genes and the environment to phenotype by comparing the concordance or correlation coefficient in monozygotic (MZ) twins with comparable values in dizygotic (DZ) twins. Hence, zygosity determination is fundamental to these studies. Questionnaires are widely used methods for zygosity classification. Although the average predictive accuracy of zygosity determination by questionnaires has reached around 95% across various studies (Rietveld et al., 2000), some studies also report less satisfactory predictive accuracy than others. Low predictive accuracy may lead to misleading conclusions, especially in small samples. Valid questionnaires are important to large epidemiologic studies as well, given their easy and low-cost nature. Thus, it is still meaningful to improve the predictive accuracy of zygosity determination questionnaires, even in modern times in which DNA genotyping technique has become more feasible.

Chinese twin samples may be particularly useful in investigating the sources of ambiguity with questionnaires and methods to reduce the ambiguity, because the accuracy of zygosity determination based on questionnaires is usually unsatisfactory in Chinese populations. For example, Gao et al. (2006) found that only 89.2% and 85.4% of adult Chinese twins were correctly classified, when using questions dealing with confusion and similarity respectively. Similarly, Lu et al. (2003) reported a predictive accuracy of 91.4% in another Chinese adult twin sample.

The main aim of the present study was to develop a unique questionnaire-based method for zygosity determination in Chinese adolescent twins, and to optimize this method by two strategies. Our first effort is to use multiple raters. Several studies found that parent-reports were useful for infant, childhood and adolescent twins (Bønnelykke et al., 1989; Chen et al., 1999; Ooki & Asaka, 2004; Price et al., 2000; Rietveld et al., 2000; Spitz et al., 1996;). For adolescents, although twins were capable of completing the questionnaires independently, answers from parents could often provide more accurate information than self-reports (Chen et al., 1999; Ooki & Asaka, 2004). In this study, we used both self-reported and parent-

Received 17 November, 2009; accepted 01 February, 2010.

Address for correspondence: Xinying Li, Address: D303, A4 Da-tun Road, Chaoyang District, Beijing, 100101, China. E-mail: lixy@psych.ac.cn

reported measures. We hypothesized that the accuracy of parent-reports would be higher than that of twins' self-reports, and the predictive accuracy would be increased by using the self-reported and parent-reported data jointly. Second, we planned to improve the predictive accuracy of zygosity questionnaires by using a 2-point response format rather than the classic 3-point one, since we speculated that the intermediate choice of each question provides more opportunities for MZ and DZ twins to respond concordantly, which would decrease the discriminating power of the questions. Adapting the 3-point scale to a 2-point one may force the twins to evaluate their situation more carefully, and to give a more deliberate answer rather than an obscure one. The validity of all questionnaires was assessed by DNA analysis of STR (short tandem repeat) variants.

Finally, cross-validation was performed in a second sample. Cross-validation is important for future use, because many statistically based prediction methods do not work well when cross-validation is attempted.

Materials and Methods

Participants and Procedure

This study is based on data from an ongoing study of adolescent behavioral and emotional disorders conducted in Beijing, China. Participants were recruited from elementary and high schools that were randomly selected from all 18 counties or districts in the Beijing municipality. For the present study, we focused on 471 pairs of same-sex twins who provided DNA samples. Among them, 255 pairs were female twins and 216 pairs were male twins. The mean age of the participants was 14.56 years (SD = 2.62). According to the results of DNA analysis, 345 pairs were MZ twins (175 female, 170 male), and 126 pairs were DZ twins (80 female, 46 male). For the cross-validation, another 382 pairs of same-sex twins were recruited. Among them, 261 pairs were MZ twins (133 female, 128 male), and 121 pairs were DZ twins (58 female, 63 male). Participants of the second sample had a mean age of 12.53 years (SD = 2.22), which was significantly younger than the mean age of the first sample (t = 16.81, p < 0.001). The reversed DZ/MZ ratio is expected in Asian populations (Bulmer, 1970; Gan et al., 2007).

Permission to conduct the research was obtained from school administrators and/or principals. Consents were obtained from the twins themselves and their parents and arrangements were made for the students to participate in their classrooms after school time. After describing the purposes of the study and explaining the procedures, trained research staff distributed the questionnaires to the participants and instructed them to complete the questionnaire independently. During the conduct of the study, teachers were present in the classrooms. Research staff were there to answer any questions that the students might have about the questionnaire. After completion, twins were instructed by our research staff to provide a saliva sample using the Oragene® DNA self-collection kit (Genotek Inc.). Questionnaires for parents were taken home by the twins and mailed back to our laboratory. The parent who spent the most time with the twins in their daily lives was asked to respond to the questionnaire. Twins and their parents were informed that they could withdraw from the study at any time. Confidentiality and anonymity of the participants' responses to the questionnaires were ensured. Institutional Review Board approved the research protocols.

Questionnaire

Each child was required to respond to two questionnaires — a 3-point one and a 2-point one. Both questionnaires consisted of 12 items which were selected, translated and adapted from zygosity questionnaires developed by Cohen et al. (1975) and Goldsmith (1991). The items were composed of two parts. The first part consisted of 6 questions dealing with physical similarity. The twins were asked if they were similar in terms of (1) hair texture, (2) height, (3) weight, (4) ear lobe, (5) overall physical appearance, and (6) how similarity changed as they grew up. The second part consisted of 6 questions dealing with confusion. The twins were asked if they were mistaken for one another by (7) parents, (8) strangers, (9) relatives, (10) close friends, (11) other classmates, and (12) teachers.

When responding to the 3-point scale, the twins made choices from 3 possible answers for each question: 1 = *No difference*, 2 = *Only a slight difference* and 3 = *Clear difference* for the first four questions; 1 = *As alike as two peas in a pod,* 2 = *As similar as typical siblings*, and 3 = *Do not look very much alike at all* for question 5; 1 = *Became more similar*, 2 = *Remained the same*, and 3 = *Became less similar* for question 6; 1 = *Often*, 2 = *Sometimes*, and 3 = *Seldom* for questions 7 to 12. When responding to 2-point scale, the twins made choices from 2 possible answers for each question: 1 = *No difference* and 2 = *Clear difference* for the first four questions; 1 = *As alike as two peas in a pod* and 2 = *As similar as typical siblings* for question 5; 1 = *No change,* and 2 = *Became quite dissimilar* for question 6; 1 = *Yes* and 2 = *No* for question 7 to 12. Parents rated the co-twin similarity and the frequency of confusion of their twin children by responding to a 3-point scale, which had similar structure and contents as those for twins.

DNA Genotyping

DNA was purified from 200 μl saliva samples using the ethanol precipitation protocol supplied with the Oragene® DNA self-collection kits. Purified DNA was redissolved in 200 μl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). PCR (polymerase chain reaction) was used to amplify the following nine STR loci: D3S1358, D13S317, D8s1179, vWA, D16S539, D18S51, D5S818, D21S11, and FGA. All of these nine loci are highly heterogeneous among the Chinese population and have been widely used for forensic testing (Fung et al., 2001; Wang et al., 2005). Oligonucleotides were synthesized as primers, and different types of fluorescent dyes were attached to each primer. The PCR reaction condition was as follows: 5.0 minutes of denaturing at 95ºC, followed by 30 cycles of 15-seconds denaturing at 94ºC, annealing at 54ºC for 30 seconds, a 30-seconds extension at 72ºC, and a final extension at 60ºC for 10 minutes. Electrophoresis was carried out to separate the PCR products. Finally a DNA sequencer (Applied Biosystems Inc.) was used to analyze the fluorescent-labeled DNA products. Twin pairs concordant for all nine loci were classified as MZ; otherwise they were classified as DZ.

Statistical Analysis

As the accuracy of questionnaires would be assessed based on the results of DNA diagnosis, the validity of DNA diagnosis itself were tested prior to the analysis of questionnaires. First, allele frequencies of the nine loci were compared to allele frequencies reported in unrelated Chinese population (Fung et al., 2001; Wang et al., 2005). Then we calculated the probability that a pair concordant at all nine loci was MZ using Bayes theorem as:

$$P(MZ\ /\ conc) = \frac{1}{1 + QPc}$$

where Q is the a priori ratio of DZ/MZ in the population and Pc is the probability that a DZ twin pair will be concordant for all markers. Based on a previous study conducted in Chinese adult twins (Gao et al., 2006), the Q value was set to 1/2.6. As the nine loci are independent, the Pc can be calculated by multiplying the probabilities that a DZ twin pair is concordant for each locus. Hence, $Pc = P_{c1}P_{c2}P_{c3}P_{c4}P_{c5}P_{c6}P_{c7}P_{c8}P_{c9}$, where $P_{c1} - P_{c9}$ are the probabilities that a DZ twin pair is concordant for a specific locus among the nine loci. We use P_{cx} to represent anyone of $P_{c1} - P_{c9}$. The calculation of P_{cx} was based on allele frequencies of that locus among the general population:

$$P_{cx} = \left[\left[1 + \sum_{i=1}^{N} p_i^{\ 2} \right]^2 + \sum_{i \neq j}^{N} \left(p_i p_j \right)^2 \right] / 4$$

where p_i and p_j are population frequencies of the i^{th} and j^{th} allele in a system containing N alleles.

Stepwise logistic regression analysis was performed to evaluate the accuracy of the questionnaires and select the most informative questions. Parsimonious models were established by using the PROC LOGISTIC program in SAS software (SAS Institute, 1993) with a significant level of .05 for entering or staying in the model. In the regression models, the zygosity classification determined by DNA genotyping was used as dependent variable (0 = *Being MZ*, 1 = *Being DZ*). The total accuracy, sensitivity and specificity of the questionnaires were calculated by specifying 'CTABLE' option in PROC LOGISTIC. As we indicated being DZ as the event, the sensitivity measured the percentage of DZ twins who were correctly classified by questionnaires, and the specificity measured the percentage of MZ twins who were correctly classified. In data analysis, twins were usually treated as individuals, except when calculating the co-twin correlation or concordance.

Results

DNA Diagnosis

All the nine loci are highly heterogeneous with 7–18 allele types for each locus. Chi-square tests showed no significant difference in allele frequencies between a general population (Fung et al., 2001; Wang et al., 2005) and our twin sample (ps > 0.29). Using the method described above, the posterior probability of being monozygotic for same-sex twins with same genotype in all the nine loci was estimated to be 99.999974%. In fact, four loci could provide sufficient information for zygosity classification. If a pair of same-sex twins were concordant in any four loci out of the above nine loci, the probability of them being MZ was greater than 99%.

Questionnaire

The internal consistency of all questionnaires was good (α = 0.760, 0.828, 0.877 for 3-point, 2-point and parent-reported questionnaires respectively). High correlations were observed in raw summed scores between co-twins, and 2-point scale led to higher co-twin correlation than 3-point one ($r = 0.727$, $p < 0.001$ for 3-point scale; $r = 0.827$, $p < 0.001$ for 2-point scale). Raw summed scores of 3-point scale correlated to that of 2-point scale ($r = 0.710$, $p < 0.001$). Parent-reports were highly correlated with twins' self-reports ($r = 0.667$, $p < 0.001$ for 3-point scale; $r = 0.769$, $p < 0.001$ for 2-point scale).

The results of the logistic regression analysis supported our hypotheses (Table 1). First, the predictive accuracy of 2-point scale was higher than that of 3-point scale. As shown in Table 1, 83.8% of the whole sample was correctly classified by using the 3-point scale (Model I), while almost 4% more were correctly predicted by adapting the 3-point scale to a 2-point one (Model II). The increase of accuracy reached statistical significance ($\chi^2 = 5.64$, $p = .018$). Second, the predictive accuracy of parent questionnaire (Model III), which is a 3-point scale per se, exceeded that of twin's 3-point questionnaire by 4.6% ($\chi^2 = 7.82$, $p = .005$), indicating that more information regarding co-twin similarity and the frequency of confusion could be gathered from twins' parents. Third, when using the self-reported and parent-reported data jointly (Model IV), another 2.9% increase of the predictive accuracy was observed on the bases of 2-point self-reported questionnaire ($\chi^2 = 4.31$, $p = .038$). Similar questions were selected by the four regression models, suggesting greater and stable distinguishing power of these items, such as 'similarity of hair texture', 'simi-

Table 1

Results of Stepwise Logistic Regression Analysis and the Predictive Accuracy of Four Parsimonious Models

Questionnaire	*N*		Logistic model	Accuracy			
	MZ	DZ	Log[p/(1-p)]=	Cut-off	Correct	Sensitivity	Specificity
I. 3-point self-reports	656	234	–8.6650+1.0012Q1+0.6675Q2+1.1310Q5+0.7233Q8 +0.5892Q12	0.60	83.8%	50.9%	95.6%
II. 2-point self-reports	656	240	–9.8314+1.5927Q1+1.8412Q5+1.6371Q8+0.6532Q11 +0.5314Q12	0.40–0.55	87.7%	68.8%–72.5%	93.3%–94.7%
III. parent-reports	622	238	–8.5133+1.1707Q1+1.0968Q5+1.1019Q8+0.7783Q12	0.50	88.4%	73.9%	93.9%
IV. 2-point self-reports and parent-reports	593	227	–9.7765+1.1303Qt1+1.0376Qt5+0.6242Qt12+0.6516 Qp1+0.4553Qp4+0.6214Qp5+1.0797Qp8	0.50	90.6%	80.2%	94.6%

Note: *N* = number of individuals; *p* = the probability of being DZ twins; Qn = question n of the questionnaire; Qtn = question n in twins' self-reports; Qpn = question n in parent-reports.

larity of height', 'frequency of confusing strangers', 'frequency of confusing teachers' and 'whether or not you are similar as two peas in a pod'.

Figure 1, Table 2 and Table 3 provide some clues for the reason why the 2-point scale is more effective than the 3-point one in zygosity classification. As shown in Figure 1, the raw scores of MZ twins and DZ twins on the 3-point and 2-point questionnaire both distributed normally. However, since the peaks of the two distributions were close to each other, a substantial overlap between them was observed. An examination of participants' responses to each question (Table 2) revealed that a large proportion of both MZ and DZ twins tended to choose the intermediate answers when responding to several questions in the 3-point scale, while this tendency was altered by using the 2-point scale (Table 3), which in turn reduced the overlapping area of raw scores of the two groups (Figure 1). Among the participants who chose the intermediate answers in the 3-point scale, MZ and DZ

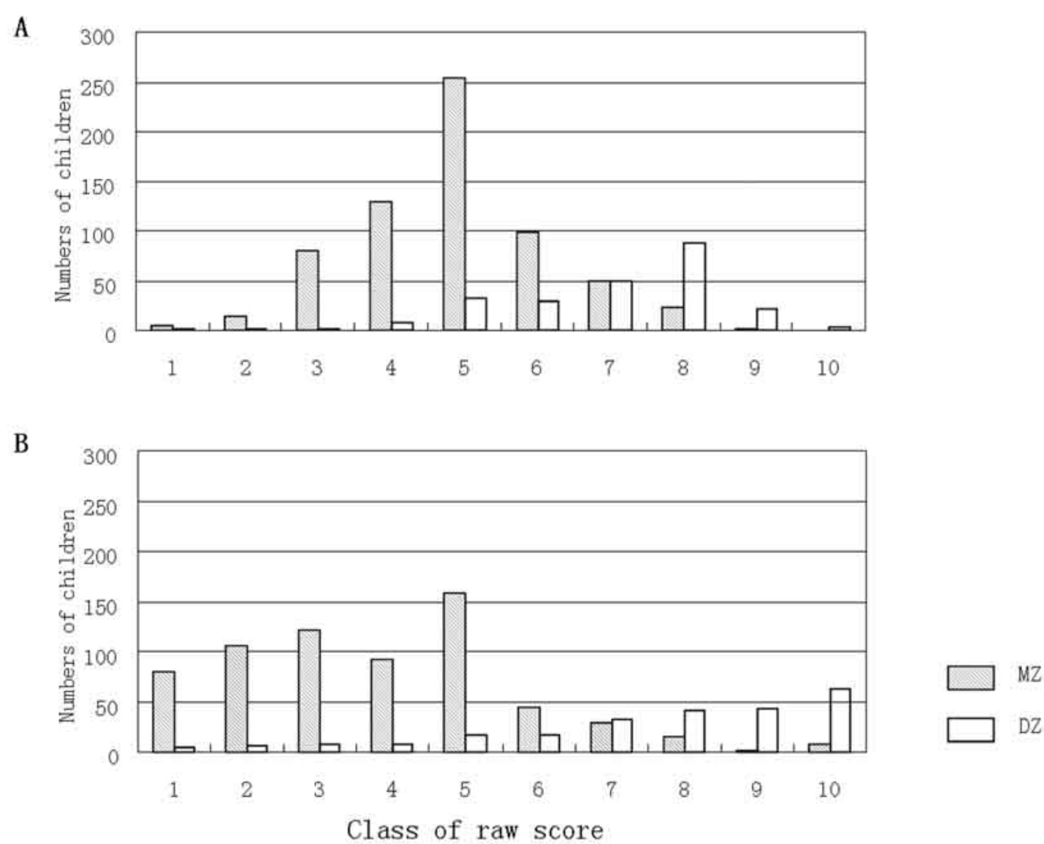

Figure 1

Raw score category of MZ and DZ twins. Participants whose raw scores fell into the first 10 percentile were counted in Class 1; participants whose raw scores fell into the second 10 percentile were counted in Class 2; and so on. (A) The distribution of twins' raw scores in the 3-point scale. (B) The distribution of twins' raw scores in the 2-point scale.

Table 2

Participants' Responses to Each Question in 3-Point and 2-Point Scales

	3-point						2-point			
	MZ (*N* = 690)			DZ (*N* = 252)			MZ (*N* = 690)		DZ (*N* = 252)	
	1	2	3	1	2	3	1	2	1	2
Similarity										
Q1: hair texture	69.6%	27.2%	2.8%	31.3%	49.2%	18.7%	90.0%	9.0%	38.5%	60.3%
Q2: height	20.0%	76.2%	3.8%	10.7%	62.3%	26.2%	80.6%	19.0%	46.4%	53.3%
Q3: weight	14.3%	71.4%	13.9%	9.9%	54.0%	34.5%	63.9%	35.7%	33.7%	65.9%
Q4: ear lobe	66.8%	30.1%	1.6%	41.7%	47.2%	9.1%	92.2%	7.2%	65.1%	33.3%
Q5: overall appearance	56.4%	33.6%	8.8%	11.1%	38.5%	48.4%	68.3%	28.6%	10.7%	88.1%
Q6: change in similarity	50.7%	38.6%	9.9%	27.8%	65.5%	4.4%	60.4%	39.0%	14.7%	82.5%
Confusion										
Q7: parents	1.2%	12.8%	85.8%	0.4%	4.4%	92.5%	6.5%	93.3%	3.2%	96.4%
Q8: strangers	70.3%	26.1%	3.3%	27.8%	40.5%	29.8%	96.2%	3.8%	51.6%	48.0%
Q9: relatives	14.3%	50.1%	35.2%	2.8%	27.4%	67.9%	56.8%	43.0%	20.6%	79.0%
Q10: close friends	7.7%	30.0%	62.0%	2.4%	12.3%	83.3%	34.6%	65.1%	11.5%	88.1%
Q11: teachers	23.2%	54.6%	21.7%	5.6%	34.1%	58.3%	70.4%	29.4%	22.6%	76.6%
Q12: classmates	22.9%	45.2%	31.6%	2.4%	12.3%	83.3%	70.9%	29.0%	19.8%	79.8%

Note: *N* = number of individuals; 0–3.1% responses were missing for each question, thus, the responses of all possible answers don not add up to 100%; lower response scores indicate greater similarity and confusion.

Table 3

Responses to the 2-Point Scale Made by the Participants Who Chose the Intermediate Option of Each Question in the 3-Point Scale

	MZ			DZ					
	N	1	2	*N*	1	2	$N_{correct}$	$N_{incorrect}$	$P_{correct}$
Q1	188	152	33	124	45	78	230	78	73.7%
Q2	526	418	105	157	82	74	492	187	72.0%
Q3	493	334	157	136	62	74	408	219	64.9%
Q4	208	176	29	119	68	49	225	97	68.8%
Q5	232	133	90	97	7	88	221	97	67.2%
Q6	266	117	148	165	13	147	264	161	61.3%
Q7	88	16	72	11	2	9	25	74	25.3%
Q8	180	163	17	102	59	42	205	76	72.7%
Q9	345	217	128	68	22	44	261	150	63.2%
Q10	207	94	112	31	8	23	117	120	49.2%
Q11	312	233	79	66	26	40	273	105	72.2%
Q12	377	268	108	86	34	50	318	142	68.7%

Note: *N* = number of individuals who chose the intermediate option of each question in the 3-point scale; one to nine responses were missing for each question; $N_{correct}$ = number of individuals choosing a correct answer for each question in the 2-point scale (i.e., MZ choosing '1' and DZ choosing '2'). $N_{incorrect}$ = number of individuals choosing an incorrect answer for each question in the 2-point scale (i.e., MZ choosing '2' and DZ choosing '1'). $P_{correct}$ = the proportion of individuals choosing a correct answer.

twins showed differential responses to the 2-point scale (Table 3). For most questions, MZ twins were more likely to choose the answers indicating greater co-twin similarity and more frequent confusion in the 2-point scale (i.e. answer '1'), while DZ twins were more likely to choose the answers indicating less co-twin similarity and less frequent confusion in the 2-point scale (i.e. answer '2'). As a result, the ambiguity caused by the uninformative intermediate answer was resolved by the 2-point scale to some extent. We calculated a $P_{correct}$ value for each question (Table 3), indicating how many twins who selected the intermediate answer in the 3-point scale made a correct choice when responding to the 2-point scale. According to the $P_{correct}$ value, not every question benefited from the 3 to 2-point transition. Interestingly, the questions benefiting more, such as Q1, Q2, Q4, Q5, Q8, Q11, and Q12, were those being selected by the logistic regression.

The results of cross-validation in the younger sample are shown in Table 4. The same accuracy was observed by using the questionnaire for parents (Model III). Slightly higher accuracy was observed by using twins' self-reports and parent-reports jointly

Table 4

Results of Cross-Validation in a Younger Sample; Calculation and Classification Were Based on Three Equations (Model II, III and IV) and Cut-Off Points Listed in Table 1

Model	*N*	Accuracy			Concordant rate
		Total	MZ	DZ	
II: self	714	84.9%	94.5%	64.2%	89.3%
III: parent	726	88.4%	93.6%	76.8%	100.0%
IV: self+parent	676	91.3%	96.6%	79.3%	98.1%

Note: *N* = number of individuals.

(Model IV). The results of cross-validation suggested good validity across different samples. Using twins' self-reports solely (Model II), however, led to a decreased accuracy than that obtained in the first sample. According to the equation and cut-off points in Model II, twin pairs in 89.3% families reported themselves as MZ or DZ concordantly. Since one parent rated his/her twin children with one questionnaire, the concordance rate of parent-reports is 100%. Using self-reports and parent-reports jointly led to a concordance rate of 98.1%.

Discussion

The present study developed and optimized a questionnaire-based method for zygosity determination in Chinese adolescent twins, and cross-validated it in an independent sample, depending upon DNA matching as the external validity criterion. Several conclusions could be drawn from our study. First, parent-reports were more informative than children's self-reports. Parent-reports showed both a higher internal consistency and predictive accuracy than self-reports. Similar results were observed in studies among Japanese (Ooki & Asaka, 2004) and Taiwanese (Chen et al., 1999) adolescents. Our study further showed that parent-reports were more stable than twins' self-reports when being tested for cross-validation. Relatively immature cognitive ability of adolescents (Blakemore & Choudhury, 2006) may account for the less accurate information provided by children themselves, and may explain why the predictive accuracy of the 3-point scale is even lower in Chinese adolescents (as in the present study) than in Chinese adults (Gao et al., 2006; Lu et al., 2003). However, children's self-reports would contribute to the increase of predictive accuracy if using the self-reported data and parent-reported data jointly, indicating benefits from multiple raters. Based on these results, we suggest future researchers should include another rater besides the twins themselves. Including an adult rater was highlighted for research among young twins.

Our second conclusion was that the 2-point scale was more useful than the 3-point scale in zygosity classification. Most previous studies used 3-point scales, and the rest used combined 2- and 3-point scales (Jackson et al., 2001; Rietveld et al., 2000). We are the first to suggest that a 3-point scale may diminish the distinguishing power of zygosity classification questionnaires and to compare the predictive accuracies of 2- and 3-point scales directly in one study. We obtained some supporting evidence for this novel idea in the present study, and adapting the 3-point scale to a 2-point one did lead to a substantial increase in predictive accuracy. Although the 2-point self-report questionnaire only led to a slight increase in the cross-validation accuracy, that was probably due to the younger age of the participants in the second sample. Thus, the 2-point scale was recommended for future studies.

Greater similarity between DZ twins may account for a large proportion of ambiguity in questionnaires in this study. Recall that DZ pairs share 50% of their genes in common *on average*, with variation around that point. As shown in Figure 1, no clear cut-off point could be decided by inspecting the distribution of raw scores, because the distribution curve of DZ twins had a long left tail which extended to the opposite extreme. Furthermore, logistic regression showed lower sensitivity than specificity (Table 1), indicating difficulties in predicting DZ twins. In cross-validation (Table 4), again, a larger proportion of DZ twins was incorrectly classified. Similar results were obtained by studies performed in Asian population, such as Japanese (Ooki & Asaka, 2004; Ooki et al., 1990), Taiwanese (Chen et al., 1999) and Chinese (Gao et al., 2006). Whereas in studies among western populations, the proportion of correctly classified MZ did not always exceed the proportion of correctly classified DZ (King et al., 1980), and a clear cut-off point in a distribution diagram could usually be decided without much difficulty (Spitz et al., 1996). We suspected that less variation in physical features, such as eye color or hair color, within Asian population may render DZ twins more similar to each other, and hence, less distinguishable. Gao et al. (2006) provided some direct evidence of this possibility. They constructed a scale which was used by a trained rater to compare the similarity of physical features between co-twins. The accuracy of this scale was only 76.9%, and the sensitivity was much lower than the specificity (45.9% vs. 88.5%), suggesting that great co-twin similarity in physical appearance exists among DZ twins as a fact, which is independent of the expertise of the raters. Since little can be done regarding the issue of less diversity in physical features among Asian people, our effort to force MZ and DZ twins to make different choices by adapting a 3-point scale to a 2-point one is certainly valuable.

We want to note that cautions should be taken when using the method established in this study in future research. Multiple raters and a 2-point scale could be used in any twin study, although the effect of enhancing predictive accuracy demands replication in other populations. The equations and cut-off points developed in this study, however, may be ethnic-specific, and may not be generalized to other populations. There is another limitation which needs to be kept in mind. Since the 2-point scale forces the respondents to make a definite choice, it may cause additional error when the respondents make too much misjudgment. Table 3 showed that a certain number of participants did misjudge when they responding to the 2-point scale (indicated by the value of $N_{incorrect}$). However, since more participants made correct choices (indicated by the value of $N_{correct}$), this degree of error should be acceptable.

Acknowledgment

One author, Dr. Xiaojia Ge, passed away during the conduct of this study. The other authors want to express their endless respect for him. This study was funded by the Chinese Academy of Sciences innovation program (KX05-066), funds for young scholars of the Institute of Psychology, Chinese Academy of Sciences (O7CX071007), and the National Natural Science Foundation of China (30600183). We thank Dr. Matt McGue for his help in composing the manuscript. We also thank T. Chen, J. Hou & J. Zhang for their work in the molecular genetic experiments.

References

Blakemore, S-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*, 296–312.

Bønnelykke, B., Hauge, M., Holm, N., Kristoffersen, K., & Gurtler, H. (1989). Evaluation of zygosity diagnosis in twin pairs below age seven by means of a mailed questionnaire. *Acta Geneticae Medicae et Gemellologiae*, *38*, 305–313.

Bulmer, M.G. (1970). *The biology of twinning in man.* Oxford, UK: Clarendon Press.

Cohen, D. J., Dibble, E., Grawe, J. M., Pollin, W. (1975). Reliably separating identical from fraternal twins. *Archives of General Psychiatry*, *32*, 1371–1375.

Fung, W. K., Ye, J., Hu, L., Zhao, X., Liu, B., Wong, D. M., & Law, M. Y. (2001). Allele frequencies for nine STR loci in Beijing Chinese. *Forensic Science International*, *121*, 207–209.

Gan, J-P., Wu, Z-H., Tu, Z-M., & Zheng, J. (2007). The comparison of twinning rates between urban and rural areas in China. *Twin Research and Human Genetics*, *10*, 633–637.

Gao, W., Li, L., Cao, W., Zhan, S., Lv, J., Qing, Y., Pang, Z., Wang, S., Chen, W., Chen, R., & Hu, Y. (2006). Determination of zygosity by questionnaire and physical features comparison in Chinese adult twins. *Twin Research and Human Genetics*, *9*, 266–271.

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.

Goldsmith, H. H. (1991). A zygosity questionnaire for young twins: A research note. *Behavior Genetics*, *21*, 257–269.

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.

King, M-C, Friedman, G. D., Lattanzio, D., Rodgers, G., Lewis, A. M., Dupuy, M. E., Williams, H. (1980). Diagnosis of twin zygosity by self-assessment and by genetic analysis. *Acta Geneticae Medicae at Gemallologiae*, *29*, 121–126.

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). Zygosity determination in epidemiology 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.

Ooki, S., Yamada, K., Asaka, A., Hayakawa, K. (1990). Zygosity diagnosis of twins by quesionnaire. *Acta Geneticae Medicae at Gemallologiae*, *39*, 109–115. (Roma).

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.

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.

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.

Wang, Z., Yu, R., Wang, F., Li, X., & Jin, T. (2005). Genetic polymorphisms of 15 STR loci in Han population form Shaanxi (NW China). *Forensic Science International*, *147*, 89–91.