Chi-squared Automatic Interaction Detection Decision Tree Analysis of Risk Factors for Infant Anemia in Beijing, China

Fang Ye[1], Zhi-Hua Chen[2], Jie Chen[1], Fang Liu[1], Yong Zhang[1], Qin-Ying Fan[1], Lin Wang[1]

[1]Department of Preventive Health Care, China-Japan Friendship Hospital, Beijing 100029, China
[2]Department of Biochemistry and Molecular Biology, China-Japan Institute of Clinical Medical Science, Beijing 100029, China

Abstract

Background: In the past decades, studies on infant anemia have mainly focused on rural areas of China. With the increasing heterogeneity of population in recent years, available information on infant anemia is inconclusive in large cities of China, especially with comparison between native residents and floating population. This population-based cross-sectional study was implemented to determine the anemic status of infants as well as the risk factors in a representative downtown area of Beijing.

Methods: As useful methods to build a predictive model, Chi-squared automatic interaction detection (CHAID) decision tree analysis and logistic regression analysis were introduced to explore risk factors of infant anemia. A total of 1091 infants aged 6–12 months together with their parents/caregivers living at Heping Avenue Subdistrict of Beijing were surveyed from January 1, 2013 to December 31, 2014.

Results: The prevalence of anemia was 12.60% with a range of 3.47%–40.00% in different subgroup characteristics. The CHAID decision tree model has demonstrated multilevel interaction among risk factors through stepwise pathways to detect anemia. Besides the three predictors identified by logistic regression model including maternal anemia during pregnancy, exclusive breastfeeding in the first 6 months, and floating population, CHAID decision tree analysis also identified the fourth risk factor, the maternal educational level, with higher overall classification accuracy and larger area below the receiver operating characteristic curve.

Conclusions: The infant anemic status in metropolis is complex and should be carefully considered by the basic health care practitioners. CHAID decision tree analysis has demonstrated a better performance in hierarchical analysis of population with great heterogeneity. Risk factors identified by this study might be meaningful in the early detection and prompt treatment of infant anemia in large cities.

Key words: Chi-squared Automatic Interaction Detection; Decision Tree Analysis; Infant Anemia; Logistic Regression Analysis

Introduction

Infant anemia is an important public health challenge worldwide. According to data of World Health Organization (WHO), the global prevalence of anemia was about 47.4% among preschool children in 2005, with more severe conditions faced by Africa and Southeast Asia.[1] The collaborative study group for "The Epidemiological Survey of Iron Deficiency in Children in China" reported a 20.5% prevalence of anemia in children based on their study from year 2000 to 2001 in China.[2] Most anemic infants within 1-year-old presented iron deficiency anemia (IDA),[1] which may impair cognitive development and, therefore, reduce their development potential.[3,4] Several studies have shown that management based on risk factors in high-risk population was effective in preventing anemia.[5-7] The major risk factors include poverty, multiple pregnancy, premature birth, low birth weight, maternal anemia, and race.[8-10] Previous studies on infant anemia focused more on the rural areas in China[11-13] while the available information in the large cities is inconclusive.

Industrialization of China has led huge changes in the composition of population in metropolis. As the capital of China, Beijing is home to a large floating population each year with one million people going in and out of downtown

Access this article online

Quick Response Code:

Website:
www.cmj.org

DOI:
10.4103/0366-6999.181955

areas every day. Chaoyang District is one of the main areas of Beijing and thus more representative of the downtown landscape. With an increasing heterogeneity of population in large cities, it is advisable to conduct hierarchical or classification analysis on the risk factors of anemia. Current research with high level of methodological rigor is limited.

In this study, we used the decision tree analyses to determine the risk factors in metropolis comparing with logistic regression models. The decision tree analyses have been developed as useful diagnostic tools in a number of clinical situations[14-20] and epidemiological investigations.[21-24] Chi-squared automatic interaction detection (CHAID) is an algorithm analysis of decision tree model based on Bonferroni test. A tree diagram generated by this method demonstrates relationship between split variables and associated related factors, which enables population subgroups with homogeneous to be revealed. Compared with logistic regression analysis, CHAID decision tree analysis enables partition of population into subgroups with different characteristics, and estimation of prevalence in each subgroup, while logistic regression analysis explores risk factors among the whole population and treats different factors equally. The combination of these two methods would contribute to better understanding of the significance of the potential factors and identify the target population.

METHODS

Participants
The research was a population-based cross-sectional study. Heping Avenue Subdistrict is one of the largest subdistricts of Chaoyang District locating between second ring to fourth ring of northeast Beijing. The population is about 130,000 with 90,000 of the native residents and 40,000 of floating population. All participants were recruited from Beijing Maternal and Child Health Network System and screened for eligibility according to the following criteria: (a) 6–12 months of age at intake between January 1, 2013 and December 31, 2014, (b) living in Heping Avenue Subdistrict for more than 3 months, and (c) parental consent. Exclusion criteria were (a) severe birth injuries (e.g., asphyxia, intracranial hemorrhage), (b) congenital diseases (e.g., phenylketonuria warranting dietary or drug treatment), (c) intestinal diseases (e.g., chronic diarrheas), (d) severe allergic diseases (e.g., severe protein allergy), and (e) acute infectious diseases 2 weeks before the study.

Study procedures
The primary health care program of children is free to residents and migrants in the local primary care hospitals. Children were evaluated by experienced examiners during routine physical examination at 1, 3, 6, 9, and 12 months, with a routine blood test at around 6–12 months of age. Parents were investigated using face-to-face interviewer-administered questionnaire. Information pertaining to pregnancy and delivery was double-checked from the Database of Beijing Maternal and Child Health Network System.

We used a systematic literature-based quantitative questionnaire to collect all the potential risk factors from the participants and their families.[2-5,8,9] The findings of the literature reviews were confirmed through interviewing with three chief physicians in the Department of Pediatrics, Hematology, and Preventive Care. This ensured that all potential risk factors associated with infant anemia were inclusive in the questionnaire. Items were modular in parental demographic profile, maternal medical history, antenatal health status, and feeding patterns. All the investigators in this study were trained under the guideline of infant anemia from the national center for maternal and child health before the survey. Each investigator collected information with a unified questionnaire. The principal investigator checked the questionnaires and called the parents for missing data every day.

The study protocol and data collection were approved by the ethical reviews boards of China-Japan Friendship Hospital. Signed informed consents were obtained before any blood test and interview from the parents/caregivers.

Diagnostic criteria and classification
According to the diagnostic criteria developed by WHO, anemia is defined by hemoglobin (Hb) <110 g/L over 6 months to 12 years, with mild anemia 90–110 g/L, moderate anemia 60–90 g/L, and severe anemia <60 g/L. In clinical practice, IDA is the most common clinical classification of microcytic hypochromic anemia, which is defined by the following parameters including mean cell volume (MCV) <80 fl, mean cell hemoglobin (MCH) <27 pg, and mean corpuscular hemoglobin concentration (MCHC) <310 g/L.

Sample size
The sample size was calculated using equation as follows:

$$n = \frac{pq}{s_p^2} = \frac{pq}{(\frac{d}{z_\alpha})^2} = \frac{z_\alpha^2 pq}{d^2}.$$

According to the result from "The Epidemiological Survey of Iron Deficiency in Children in China"[2], the prevalence of children in China was 20.5% ($P = 0.205$, q = 1-0.205). With a hypothesizing $\alpha = 0.05$ (two-sided), $z_\alpha = 1.96$, and $d = 0.15 \times p$.

Calculating the sample size:

$$n_0 = \frac{1.96 \times 1.96 \times 0.205 \times (1 - 0.205)}{(0.15 \times 0.205)^2} = 662.$$

The minimum sample size was inflated to compensate for nonresponses and desired hierarchical analyses:

$$n_1 = n_0 \times 1.6 = 1059.$$

The proposed sample size was 1059 participants.

Statistical analysis

EpiData 3.0 software (Epidata Association, Ddense, Denmark) was introduced to establish a database and dual data input, followed by data correction. SPSS version 18.0 (SPSS Inc., Chicago, Illinois, USA) was used for descriptive analysis, univariate analysis, CHAID decision tree analysis, and logistic regression analysis. Hb, the core variable of anemia, and other potential risk factors were included in descriptive analysis. The prevalence of infant anemia with potential risk factors was compared by univariate analysis. A CHAID decision tree analysis was applied to identify potential factors and determine their relationships with infant anemia following univariate analysis. In CHAID analysis, anemia was the target variable and risk factors were explanatory variables. The Pearson's Chi-squared test and maximum likelihood classification were used to compare different categorical variables, which were classified into a binary or more series by the most significant predictor. Then as in the first step, cases in each subgroup are further partitioned by the second most significant predictor. The analysis continues in this manner until the last significant predictors were done.[25] A multivariate logistic regression with multicollinearity test was conducted as a comparison to CHAID analysis. The overall accuracy was expressed as percentage in both analyses. The significance level for node splitting of decision tree in CHAID and in logistic regression analysis was $P < 0.05$.

RESULTS

Demographic profile

A total of 1091 infants aged 6–12 months were included in this study after excluding two cases of severe birth asphyxia, three with congenital diseases, one with chronic diarrhea, two with history of severe chronic diseases, and 12 with acute infectious diseases 2 weeks before the latest medical examination.

Table 1 lists demographic information as well as Hb level. The mean value of Hb was 119.57 ± 8.05 g/L (range 52.00–143.00 g/L). There were 137 (12.60%) cases of microcytic hypochromic anemia as indicated by values of MCV <80 fl, MCH <27 pg, and MCHC <310 g/L, including one case of severe anemia, six of moderate anemia, and 130 of mild anemia. The case of severe anemia was a preterm of 32 weeks of gestation with very low birth weight of 1780 g, whose parents were floating population from rural areas in Gansu Province. After discharge from the Department of Pediatrics of the hospital, the infant was introduced to breastfeeding without attending regular health examination and any nutritional supplements. According to the latest published "Indicators for assessing infant and young child feeding practices" by WHO, exclusive breastfeeding (EBF) is defined as infant receives breast milk (including expressed breast milk or breast milk from a wet nurse) and allows the infant to receive oral rehydration salts, drops, syrups (vitamins, minerals, and medicines), but nothing else. Formula feeding is defined as infant receives formula or nonhuman milk only. Mixed

Table 1: Demographic profile of participants (*n* = 1091)

Characteristics	Values	Proportion (%)
Hemoglobin* (g/L)	119.57 ± 8.05	
<59 (severe)	1	0.1
60–89 (moderate)	6	0.5
90–109 (mild)	130	11.9
110–129	902	82.7
>130	52	4.8
Gender		
Male	544	49.9
Female	547	50.1
Gestational age at birth (weeks)		
<37	51	4.7
37–41	1030	94.4
≥42	10	0.9
Birth weight (g)		
<2500	51	4.7
2500–4000	977	89.5
>4000	63	5.8
Obesity		
Yes	33	3.0
No	1058	97.0
Delivery		
Vaginal delivery	583	53.4
Cesarean section	508	46.6
Twins		
Yes	16	1.5
No	1075	98.5
Residence		
Floating population	341	31.3
Beijing residents	750	68.7
Maternal anemia		
Yes	65	6.0
No	1026	94.0
Feeding patterns in first 6 months		
Exclusive breastfeeding	836	76.6
Mixed feeding	218	20.0
Formula feeding	37	3.4
Mother's age (years)	30.4 ± 3.7	
Mother's educational level		
College, university or above	935	85.8
High school or under	156	14.2

Values are *n* or mean ± SD. *The MCV, MCH, and MCHC were also evaluated in patients diagnosed as anemia. SD: Standard deviation; MCV: Mean cell volume; MCH: Mean cell hemoglobin; MCHC: Mean corpuscular hemoglobin concentration.

feeding is defined as combination of breast and formula feeding.[26]

In our cases, microcytic hypochromic anemia may not only be IDA but also related to other factors (such as thalassemia) in rare cases. To estimate the possibility caused by other factors, we added iron-rich complementary foods such as yolk and liver paste to the diets of 130 infants with mild anemia and administered iron therapy (ferrous sulfate 2–3 $mg \cdot kg^{-1} \cdot d^{-1}$) to seven cases of moderate or severe anemia. Two months later, all 137 anemic infants were reexamined for Hb. The result revealed elevation of Hb, indicating all the anemic infants were IDA.

Univariate analysis of risk factors for anemia

As shown in Table 2, 40.0% of babies borned in anemic pregnancy were anemic, which was significantly higher than that borned in nonanemic pregnancy ($P < 0.001$). A remarkably higher proportion of anemia was observed in infants whose both parents were floating population when comparing with those who were Beijing residents with one or both parents ($P = 0.009$). During the first 6 months of life, more infants of EBF developed anemia when comparing with mixed feeding or formula ($P = 0.003$). No differences in the rest of the factors were detected between anemic and nonanemic infants (all $P > 0.05$).

Chi-squared automatic interaction detection decision tree analysis of anemia

Figure 1 shows the CHAID decision tree analysis for anemia of infants aged 6–12 months. Each node contains three statistical values (category, %, n): in addition to node number, "n" stands for anemic or nonanemic population in this particular category, "%" is the percentage of anemic or nonanemic population.

As shown in Figure 1, there were four main variables affecting infant anemia: the first variable was split on maternal anemia, and followed by EBF in first 6 months, floating population, and mother's educational level with significance of $P < 0.05$. Other variables did not reach significance of 0.05 and were not included in the model such as gender, gestational age, birth weight, twins, and delivery type.

The four selected variables were used for grouping in the decision tree model. The model includes a total of 11 nodes with 6 terminal nodes (number 2, 5, 6, 7, 9, and 10). In each category, the population of anemia differed significantly (all $P < 0.05$). The estimated error of risk in the model was 0.126, and the standard error was 0.010.

Logistic regression analysis of infant anemia

Table 3 demonstrates the results from multivariate logistic regression analysis of risk factors of infant anemia. The model had no violation of multicollinearity with the model reaching significance, with the first variable split on maternal anemia, followed by EBF in first 6 months, and floating
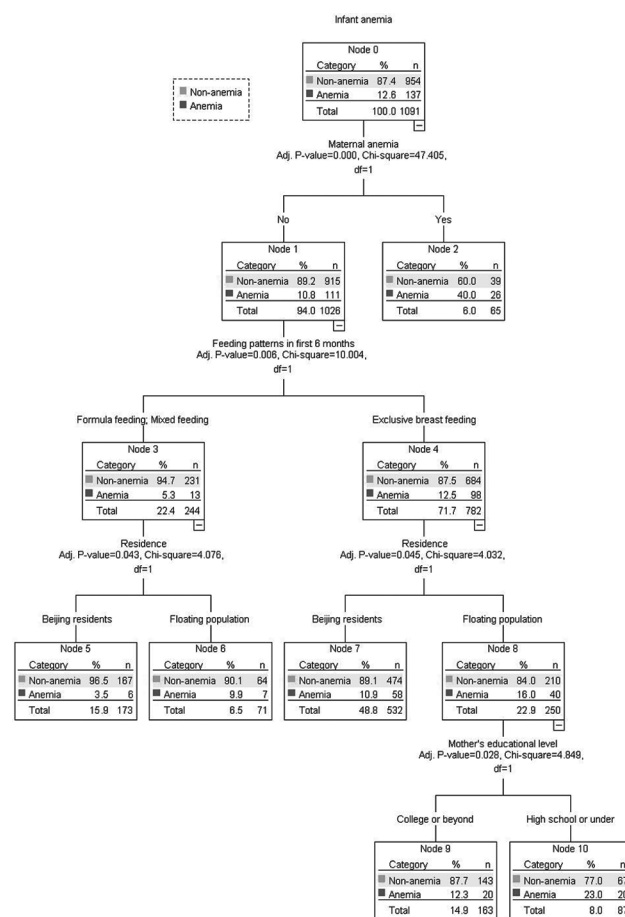
Table 2: Univariate analysis of influencing factors of infant anemia

Factors	Anemia, n (%)	Nonanemia, n (%)	χ^2	P
Gender				
Male	78 (14.3)	466 (85.7)	3.134	0.077
Female	59 (10.8)	488 (89.2)		
Gestational age (weeks)				
<37	9 (17.6)	42 (82.4)	2.366	0.286
37–41	126 (12.2)	904 (87.8)		
≥42	2 (20.0)	8 (80.0)		
Birth weight (g)				
<2500	10 (19.6)	41 (80.4)	3.393	0.183
2500–4000	127 (12.2)	850 (87.0)		
>4000	5 (9.5)	58 (92.1)		
Twins				
Yes	0	16 (100)	2.332	0.127
No	137 (12.7)	938 (87.3)		
Delivery				
Vaginal delivery	66 (11.3)	517 (88.7)	1.743	0.187
Cesarean section	71 (14.0)	437 (86.0)		
Maternal anemia				
Yes	26 (40.0)	39 (60.0)	47.405	<0.001
No	111 (10.8)	915 (89.2)		
Residence				
Beijing residents	81 (10.8)	669 (89.2)	6.748	0.009
Floating population	56 (16.4)	285 (83.6)		
Mother's age (years)				
≥35	18 (13.4)	116 (86.6)	0.107	0.744
<35	119 (12.4)	838 (87.6)		
Mother's educational level				
College, university or above	112 (12.0)	823 (88.0)	1.994	0.158
High school or under	25 (16.0)	131 (84.0)		
Feeding patterns in first 6 months				
EBF	120 (14.4)	716 (85.6)	11.196	0.003
Mixed feeding	14 (6.4)	204 (93.6)		
Bottle feeding	3 (8.1)	34 (91.9)		

EBF: Exclusive breastfeeding.

Figure 1: A CHAID decision classification tree analysis to identify the factors of infant anemia in Heping Avenue Subdistrict of Beijing. CHAID: Chi-squared automatic interaction detection.

population with significance of $P < 0.05$. Other variables such as mother's educational level, gender, gestational age, birth weight, twins, and delivery type did not reach significance of 0.05 and, therefore, were not singled out in the model.

Results from the comparison of logistic regression analysis and Chi-squared automatic interaction detection decision tree analysis

The overall classification accuracy of CHAID analysis was higher than that of logistic regression model (88.8% vs. 87.2%). ROC curve stands for receiver operating characteristic or relative operating characteristic. It is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. The area under the curve is a measure of test accuracy. As shown in Figure 2, the area below ROC curve of CHAID decision tree was larger than that of logistic regression analysis with marginal significance (0.722 vs. 0.656, $P = 0.062$), both of which were away from reference line.

DISCUSSION

To the best of our knowledge, the population-based cross-sectional study using CHAID analysis to focus on the most vulnerable population of infant anemia was rarely reported in metropolis of China. A total of 1091 infants aged

Table 3: Multivariate logistic regression analysis of influencing factors of infant anemia

Variables	B	SE	P	OR	95% CI for Exp(B)
Maternal anemia	1.699	0.277	<0.001	5.468	3.178–9.409
Exclusively breastfed in first 6 months	0.811	0.274	0.003	2.249	1.314–3.849
Floating population	0.499	0.193	0.010	1.647	1.128–2.405

B: Coefficient; *SE*: Standard error; *OR*: Odd ratio, which equals to the exponentiation of the *B* coefficient; 95% *CI* for Exp (*B*): 95% confidence interval of the exponentiation of the *B* coefficient.
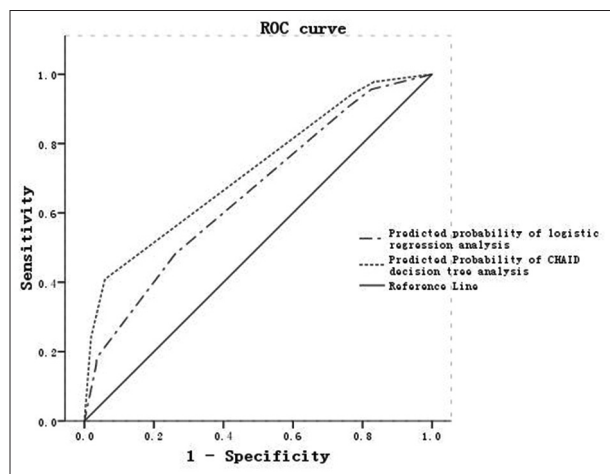

Figure 2: ROC curves of CHAID and logistic regression analysis. ROC: Receiver operating characteristic; CHAID: Chi-squared automatic interaction detection.

6–12 months living at Heping Avenue Subdistrict of Beijing were surveyed in this study. The results indicated that the prevalence of anemia was 12.6% although it was lower than China's or global level,[2] one-eighth of the participants were anemic in downtown Beijing was unexpected. The CHAID decision tree model has demonstrated multilevel interaction among risk factors through stepwise pathways to detect anemia. Risk factors were discussed by hierarchical order as follows.

Our study detected that maternal anemia was the first layer of predictor in the decision tree, indicating that it was the primary and most important risk factor for infant anemia. Up to 40.0% of infants suffering from maternal anemia was in accordance with previous reports.[11,12,27,28] These results indicated that infants who suffered from maternal anemia should be identified as the highest risk group. However, there are no preventive guidelines specifically for early blood examination and iron supplementation for these infants in China so far. In clinical practice, obstetricians focus more on severe conditions such as pregnancy-induced hypertension, diabetes mellitus, and placenta previa and less on anemia. We, therefore, recommend increasing education and perceived knowledge on the anemic pregnancy for obstetricians. In addition, infants born to anemic mothers should be singled out and specially managed by the pediatrician in primary health care hospitals.

Our results illustrated that EBF was another important risk factor for those infants not suffering from maternal anemia. According to surveys conducted in Shaanxi, Yinchuan, and Beijing, EBF infants have a higher incidence rate of anemia than those of mixed and formula feeding, consistent with our results.[11,12,29]

During the 3rd trimester of pregnancy, fetus derives iron nutrition from the mother to meet most of the demand for the first 4–6 months of age. After that, along with the infant rapid growth, the iron reserves are exhausted which results in a rapidly elevated iron demand from 0.27 mg/d on 0–6 months to 11 mg/d on 7–12 months.[30] If EBF is still strictly obeyed in this period without any intake of iron supplements, the infant may develop IDA. Hence, the lack of iron supplements after 4 months of EBF has been listed by the American Academy of Pediatrics as one of the risk factors for infant anemia.[30] A daily supplementation of oral iron of 1 mg/kg is recommended for EBF infants in the first 6 months of life until iron-fortified complementary foods (such as iron-fortified rice flour) are given.[31]

Based on the WHO statement released in 2001, infant feeding guideline in China recommends EBF for the first 6 months, without introducing any nutritious complementary foods.[32] However, as a global imperative, the recommended length of EBF and also the age of the introduction of complementary foods remained controversial and debated in literature.[33,34] In the past few years, the suggestion of EBF in the first 6 months, which is widely propagated in China, is misunderstood. Some parents do not recognize the significance and therefore refuse to add any complementary

foods or iron supplement in the first 6 months, especially the mother with higher educational level. This has become an important reason of infant anemia and, thus, we should interpret EBF and advise the parents correctly.

In contrast to previous studies, our study shows slightly but not significantly higher prevalence of anemia for premature and low birth weight infants than full-term ones. A total of 51 preterm infants were administered preventive treatment in our study. The increased perception and preventive treatment of iron may play an important role in it. After further analysis, we found lower proportions of EBF in premature and low birth weight infants than full-term infants (58.8% vs. 81.2%), which may also contribute to this positive result.

Residential status and maternal educational level are the third and fourth layers of CHAID decision tree, respectively, indicating that floating population and low maternal educational level were both risk factors for these anemic infants. The extremely severe anemia (Hb = 52 g/L) was occurred in a preterm baby of floating population with maternal education of middle school also verified the effect of these two factors. A few of floating population does not undergo routine physical examination comparing with resident population (75.2% vs. 92.7%, $P < 0.05$), indicating significantly less focus on infant health. Although both residents and floating population enjoy the same free social welfare benefits, the latter still fails to enjoy them, probably due to work-related pressures. Most poorly educated floating population work in labor-intensive industry, such as construction workers, petty dealers. The financial burden and job intensity lead to neglect of infant health. Poor health education on child care also results in ignorance of infant health issues.

For the comparison of the two statistical methods, logistic regression analysis predicts the probabilities of dependent based on one or more independents using cumulative logistic distribution. It treats all independents equally, ignoring the hierarchical relation among them. CHAID analysis focuses on hierarchical relation by crossed classification to generate subgroups with different characteristics and deals with interaction effect among independents by generating an image that is easy to interpret.

Similar with results from logistic regression analysis, CHAID identified maternal anemia, EBF feeding pattern, and residential status as major risk factors of infant anemia. In addition, CHAID analysis further identified the maternal educational level in floating population as a risk factor in the fourth layer of the decision tree model. This risk factor was not detected in logistic regression analysis because of higher proportion of mothers in floating population was tended to be with lower education level, which cover the fact. That is to say, lower maternal educational level was not a predictor of anemia in all participants, but only in those exclusive breastfed floating population whose mother was not suffered from anemia during pregnancy. Compared with logistic regression analysis, CHAID decision tree analysis was closer to reality and provided an effective approach to detect target population for further intervention.

In this population-based study, we evaluated the anemic status of infants aged 6–12 months whose parents either were Beijing residents or belonging to floating population. We analyzed different characteristics of anemic infants in detail using decision tree classification of CHAID analysis and logistic regression analysis. In addition to the high-risk factors such as maternal anemia which are consistent with previous studies, pediatricians should focus on clearing EBF-related myths as well as promote evidence-based screening and education of high-risk floating population. The pathway derived from this model will shed light on the prediction of the high-risk population for anemia prevention and promote the update of guidelines on infant anemia.

In further studies, following up of both anemic and health infant cohorts are strongly suggested to explore long-term effect of identified risk factors on their physical growth, cognitive level, and motor development. In addition, the current study was a pilot study to explore the potential predictors of infant anemia in metropolis. Therefore, the relatively smaller sample size may affect the stability of results from CHAID algorithm which requires large sample sizes across the country. Further study can enroll more participants from different regions to improve the results' stability and representatives.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. McLean E, Cogswell M, Egli I, Wojdyla D, de Benoist B. Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005. Public Health Nutr 2009;12:444-54. doi: 10.1017/S1368980008002401.
2. China Iron Deficiency Epidemiologic Study Group. Study on iron deficiency of children aged over 7 months to 7 years old in China (in Chinese). Chin J Pediatr 2004;42:886-91. doi: 10.3760/j.issn:0578-1310.2004.12.003.
3. Yadav D, Chandra J. Iron deficiency: Beyond anemia. Indian J Pediatr 2011;78:65-72. doi: 10.1007/s12098-010-0129-7.
4. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, *et al*. Child development: Risk factors for adverse outcomes in developing countries. Lancet 2007;369:145-57. doi: 10.1016/S0140-6736(07)60076-2.
5. Imdad A, Bhutta ZA. Intervention strategies to address multiple micronutrient deficiencies in pregnancy and early childhood. Nestle Nutr Inst Workshop Ser 2012;70:61-73. doi: 10.1159/000337441.

6. De-Regil LM, Suchdev PS, Vist GE, Walleser S, Peña-Rosas JP. Home fortification of foods with multiple micronutrient powders for health and nutrition in children under two years of age (Review). Evid Based Child Health 2013;8:112-201. doi: 10.1002/14651858. CD008959.pub2.

7. Lazzerini M, Rubert L, Pani P. Specially formulated foods for treating children with moderate acute malnutrition in low- and middle-income countries. Cochrane Database Syst Rev 2013;6:CD009584. doi: 10.1002/14651858.CD009584.pub2.

8. Meinzen-Derr JK, Guerrero ML, Altaye M, Ortega-Gallegos H, Ruiz-Palacios GM, Morrow AL. Risk of infant anemia is associated with exclusive breast-feeding and maternal anemia in a Mexican cohort. J Nutr 2006;136:452-8. doi: 10.2105/AJPH.83.8.1130.

9. Brotanek JM, Gosz J, Weitzman M, Flores G. Iron deficiency in early childhood in the United States: Risk factors and racial/ethnic disparities. Pediatrics 2007;120:568-75. doi: 10.1542/peds.2007-0572.

10. Kilbride J, Baker TG, Parapia LA, Khoury SA, Shuqaidef SW, Jerwood D. Anaemia during pregnancy as a risk factor for iron-deficiency anaemia in infancy: A case-control study in Jordan. Int J Epidemiol 1999;28:461-8. doi: 10.1093/ije/28.3.461.

11. Yang W, Li X, Li Y, Zhang S, Liu L, Wang X, *et al.* Anemia, malnutrition and their correlations with socio-demographic characteristics and feeding practices among infants aged 0-18 months in rural areas of Shaanxi province in Northwestern China: A cross-sectional study. BMC Public Health 2012;12:1127. doi: 10.1186/1471-2458-12-1127.

12. Ma JF, Zhang H, Wang BZ. Analysis of risk factors of iron deficiency anemia for 6-month-old infant in Yinchuan (in Chinese). Ningxia Med J 2014;36:437-9. doi: 10.13621/j.1001-5949.2014.05.0437.

13. Lu LM, Zhao WJ, Dong WH. Correlation analysis of iron-deficiency in late pregnancy and infant anemia (in Chinese). Chin Prim Health Care 2009;23:55-6. doi: 10.3969/j.issn.1001-568X.2009.08.025.

14. Buntinx F, Truyen J, Embrechts P, Moreel G, Peeters R. Evaluating patients with chest pain using classification and regression trees. Fam Pract 1992;9:149-53. doi: 10.1093/fampra/9.2.149.

15. Tsien CL, Fraser HS, Long WJ, Kennedy RL. Using classification tree and logistic regression methods to diagnose myocardial infarction. Stud Health Technol inform 1998;52 Pt 1:493-7.

16. Goldman L, Weinberg M, Weisberg M, Olshen R, Cook EF, Sargent RK, *et al.* A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. N Engl J Med 1982;307:588-96. doi: 10.1056/NEJM198209023071004.

17. Selker HP, Griffith JL, Patil S, Long WJ, D'Agostino RB. A comparison of performance of mathematical predictive methods for medical diagnosis: Identifying acute cardiac ischemia among emergency department patients. J Investig Med 1995;43:468-76.

18. Stewart PW, Stamm JW. Classification tree prediction models for dental caries from clinical, microbiological, and interview data. J Dent Res 1991;70:1239-51. doi: 10.1177/00220345910700090301.

19. Lieu TA, Quesenberry CP, Sorel ME, Mendoza GR, Leong AB. Computer-based models to identify high-risk children with asthma.

20. Herman WH, Smith PJ, Thompson TJ, Engelgau MM, Aubert RE. A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes. Diabetes Care 1995;18:382-7. doi: 10.2337/diacare.18.3.382.

21. Nelson LM, Bloch DA, Longstreth WT Jr., Shi H. Recursive partitioning for the identification of disease risk subgroups: A case-control study of subarachnoid hemorrhage. J Clin Epidemiol 1998;51:199-209. doi: 10.1016/S0895-4356(97)00268-0.

22. Zhang H, Holford T, Bracken MB. A tree-based method of analysis for prospective studies. Stat Med 1996;15:37-49. doi: 10.1002/(SICI)1097-0258(19960115)15:1<37::AID-SIM144>3.0.CO;2-0.

23. Carmelli D, Zhang H, Swan GE. Obesity and 33-year follow-up for coronary heart disease and cancer mortality. Epidemiology 1997;8:378-83. doi: 10.1097/00001648-199707000-00005.

24. Zhang H, Bracken MB. Tree-based risk factor analysis of preterm delivery and small-for-gestational-age birth. Am J Epidemiol 1995;141:70-8.

25. Kass GV. An exploratory technique for investigating large quantities of categorical data. Appl Stat 1980;29:119-27. doi: 10.2307/2986296.

26. World Health Organization. Indicators for Assessing Infant and Young Child Feeding Practices. Available from: http://www.apps.who.int/iris/bitstream/10665/43895/1/9789241596664_eng.pdf. [Last accessed on 2015 Mar 24].

27. Chantry CJ, Howard CR, Auinger P. Full breastfeeding duration and risk for iron deficiency in U.S. infants. Breastfeed Med 2007;2:63-73. doi: 10.1089/bfm.2007.0002.

28. De Pee S, Bloem MW, Sari M, Kiess L, Yip R, Kosen S. The high prevalence of low hemoglobin concentration among Indonesian infants aged 3-5 months is related to maternal anemia. J Nutr 2002;132:2215-21.

29. Gong YH, Ji CY, Zheng XX, Shan JP, Hou R. Correlation of 4-month infant feeding modes with their growth and iron status in Beijing. Chin Med J 2008;121:392-8.

30. Baker RD, Greer FR; Committee on Nutrition American Academy of Pediatrics. Diagnosis and prevention of iron deficiency and iron-deficiency anemia in infants and young children (0-3 years of age). Pediatrics 2010;126:1040-50. doi: 10.1542/peds.2010-2576.

31. Friel JK, Aziz K, Andrews WL, Harding SV, Courage ML, Adams RJ. A double-masked, randomized control trial of iron supplementation in early infancy in healthy term breast-fed infants. J Pediatr 2003;143:582-6. doi: 10.1067/S0022-3476(03)00301-9.

32. National Health and Family Planning Commission of the People's Republic of China. Infant Feeding Guidelines; 2012. Available from: http://www.moh.gov.cn/cmsresources/mohfybjysqwss/cmsrsdocument/doc14756.doc. [Last accessed on 2015 Mar 14].

33. Forsyth JS. Policy and pragmatism in breast feeding. Arch Dis Child 2011;96:909-10. doi: 10.1136/adc.2011.215376.

34. Fewtrell M, Wilson DC, Booth I, Lucas A. Six months of exclusive breast feeding: How good is the evidence? Br Med J 2011;342:c5955. doi: 10.1136/bmj.c5955.