ARTICLE

Noninherited Factors in Fetal Congenital Heart Diseases Based on Bayesian Network: A Large Multicenter Study

Yanping Ruan[1,#], Xiangyu Liu[2,#], Haogang Zhu[3,*], Yijie Lu[3], Xiaowei Liu[1], Jiancheng Han[1], Lin Sun[1], Ye Zhang[1], Xiaoyan Gu[1], Ying Zhao[1], Lei Li[2], Suzhen Ran[4], Jingli Chen[5], Qiong Yu[6], Yan Xu[7], Hongmei Xia[8] and Yihua He[1,*]

[1]Department of Echocardiography, Maternal-Fetal Medicine Research Consultation Center, Beijing Anzhen Hospital, Capital Medical University, Beijing, 100029, China

[2]School of Biological Science and Medical Engineering, Beihang University, Beijing, 100083, China

[3]State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing, 100191, China

[4]Department of Ultrasound, Chongqing Health Center for Women and Children, Chongqing, 400013, China

[5]Department of Ultrasound, Urumqi Maternal and Child Health Hospital, Urumqi, 830001, China

[6]Department of Ultrasound, Pingxiang Maternal and Child Care, Pingxiang, 337055, China

[7]Department of Ultrasound, People's Hospital of Rizhao, Rizhao, 276800, China

[8]Department of Ultrasound, Xinqiao Hospital Army Medical University, Chongqing, 400038, China

[*]Corresponding Authors: Yihua He. Email: heyihuaecho@hotmail.com; Haogang Zhu. Email: haogangzhu@buaa.edu.cn

[#]These authors contributed equally to the manuscript

Received: 07 February 2021   Accepted: 12 April 2021

ABSTRACT

Background: Current studies have confirmed that fetal congenital heart diseases (CHDs) are caused by various factors. However, the quantitative risk of CHD is not clear given the combined effects of multiple factors. **Objective:** This cross-sectional study aimed to detect associated factors of fetal CHD using a Bayesian network in a large sample and quantitatively analyze relative risk ratios (RRs). **Methods:** Pregnant women who underwent fetal echocardiography (N = 16,086 including 3,312 with CHD fetuses) were analyzed. Twenty-six maternal and fetal factors were obtained. A Bayesian network is constructed based on all variables through structural learning and parameter learning methods to find the environmental factors that directly and indirectly associated with outcome, and the probability of fetal CHD in the two groups is predicted through a junction tree reasoning algorithm, so as to obtain RR for fetal CHD under different exposure factor combinations. Taking into account the effect of gestational week on the accuracy of model prediction, we conducted sensitivity analysis on gestational week groups. **Results:** The single-factor analysis showed that the RRs for the numbers of births, spontaneous abortions, and parental smoking were 1.50, 1.38, and 1.11 ($P < 0.001$), respectively. The risk gradually increased with the synergistic effect of ranging from one to more environmental factors above. The risk was higher among subjects with five synergistic factors, including the number of births, upper respiratory tract infection during early pregnancy, anemia, and mental stress as well as a history of spontaneous abortions or parental smoking, than in those with less than 5 factors (RR = 2.62 or 2.28, $P < 0.001$). This result was consistent across the participants grouped by GWs. **Conclusion:** We identified six factors that were directly associated with fetal CHD. A higher number of these factors led to a higher risk of CHD. These findings suggest that it is important to strengthen healthcare and prenatal counseling for women with these factors.

Abbreviations

CHD:	Congenital heart disease
BN:	Bayesian network
RR:	Risk ratio
AHA:	American heart association
ASE:	American Society of Echocardiography
ISUOG:	International Society of Ultrasound in Obstetrics and Gynecology
LVOT:	Left and right ventricular outflow tract
RVOT:	Right ventricular outflow tract
3V:	Three-vessel
3VT:	Three vessels and trachea
IQR:	Interquartile range
DAG:	Directed acyclic graph
CPT:	Conditional probability table
MLE:	Maximum likelihood estimation
AUC:	Area under the curve
GWs:	Gestational weeks
MWST:	Maximum weight spanning tree
K2:	With two random initializations
K2+T:	K2 with MWST initialization
K2-T:	K2 with inverse MWST initialization
GS:	Starting from an empty structure
GS+T:	GS starting from a MWST-initialized structure
GES:	Greedy search in the space of equivalent classes

1 Introduction

Congenital heart disease (CHD) is a large, rapidly emerging global problem in child health; in the Global, regional, and national burden of CHD Study 2017, the global prevalence of CHD is estimated to be nearly 1.8‰ [1]. It varies depending on the type of defect [2]. CHD is one of the most common birth defects in the world, and one of the leading causes of morbidity and mortality in the perinatal and infant periods [3,4]. The incidence of fetal CHD cannot be reasonably estimated, but it is often misestimated because of prenatal detection rates and intrauterine loss. A study of CHD performed in 111,225 births in Belgium reported an incidence of 8.3‰ among live and stillborn infants with more than 26 weeks of gestation and no chromosomal abnormalities [5]. Many factors are associated with an increased risk of CHD in the fetus, and they can be divided into genetic [6] and maternal or fetal factors, both of which are addressed in detail for their associations with various risks in the Scientific Statement from the American Heart Association (AHA) [7].

However, the findings presented in previous studies are usually based on logistic regression analysis and the premise that the factors are independent with each other. However, CHD is a multifactorial complex disease. Current research indicates that the interactions between factors and the synergistic effects of multiple factors on the risk of CHD are difficult to consider and accurately estimate.

Therefore, this study aims to establish an optimal Bayesian network (BN) model according to our large maternal-fetal dataset, detect factors that influence fetal CHD and quantify the extent of their correlations. This method mainly shows the causal relationships among variables by performing probabilistic reasoning and reflects the potential relationships among multiple factors [8]. Moreover, an accurate estimation of the synergistic effects of multifactor combinations on fetal CHD is needed in order to manage and control the factors associated with fetal CHD, thus further reducing its incidence.

2 Materials and Methods

2.1 Participants and Study Design

In all, 16,086 pregnant women who underwent fetal echocardiography from June 2010 to June 2017 at our center and other participating centers were consecutively recruited to reduce the potential bias partially. All data are kept in our center's maternal-fetal medicine database. Data from our center accounted for the main part of the dataset. Fetal ages ranged from 16 to 39 gestational weeks (GWs) and were calculated from the last menstrual period.

Enrollment criteria included (1) relatively complete data, and (2) fetal heart examination and diagnosis of fetal CHD meeting the specifications for the diagnosis and treatment of fetal cardiac disease released by AHA [7]. The exclusion criteria were (1) subjects with more than 50% missing variables and (2) variables with more than 50% missing data. This cross-sectional study was approved by the Institutional Review Board of Beijing Anzhen Hospital, Capital Medical University, and written informed consent was signed by all participants.

2.2 Data Collection and Variable Assignment

The data and fetal echocardiography images obtained from the participating institutions were analyzed in our center. They were exported from our maternal-fetal database together with our center's data. We acquired the following information through questionnaires while patients were waiting to be examined: maternal factors included age, comorbidities (diabetes, upper respiratory infection during early pregnancy, anemia, connective tissue diseases and thyroid disease), medication exposure, history of induced labor or spontaneous abortion or CHD, consanguineous marriage, occupation and bad habits of subjects and spouses, radioactive substances exposure, and mental stress during the early pregnancy. The diagnostic cut-off for Diabetes was set at 5.3–10.0–8.6 mmol/L in 75 g OGTT [9]. Anemia was defined as hemoglobin (g/dL) and hematocrit (percentage) levels below 11 g/dL and 33%, respectively, in the first trimester; 10.5 g/dL and 32%, respectively, in the second trimester; and 11 g/dL and 33%, respectively, in the third trimester [10]. The pregnant women were asked whether they experienced various stressful events during early pregnancy, such as sickness or death for close persons, family conflict, heavy work pressure and not getting along with colleagues. And they were also asked to indicate their subjective responses to these stressors, such as positive or negative. Maternal mental stress is defined as that the pregnant women suffered from stressful events and exhibited negative attitude. The fetal factors included GWs, single or twin fetus pregnancy, fetal hydrops, and fetal arrhythmia. In addition, the subjects were divided into two groups, a normal fetus group and a CHD fetus group. According to the enrollment and exclusion criteria, 16,086 records and 26 variables were included in the final analysis. The variables and their assignment are presented in Tab. 1.

2.3 Diagnosis of Fetal CHD

The diagnosis of fetal CHD was based on fetal echocardiography using a Voluson E8-RAB4-8 machine equipped with a 2- to 8-MHz transducer (GE Healthcare, Little Chalfont, United Kingdom), an Aloka 10, UST-9130 equipped with a 3- to 6-MHz transducer (Aloka, Tokyo, Japan), or a Philips IU-22, C5-1 equipped with a 1- or 5-MHz transducer (Philips Healthcare, Bothell, WA, USA). The acquisition

of fetal echocardiographic images was performed according to the guidelines and standards of the AHA 7 and the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) [11].

Table 1: The variables and their assignment in the study

N of Variables	Variables	Value	Group details
1	Maternal age (years)	1	<20
1		2	$20 \leq age < 35$
1		3	≥35
2	Induced labor	1	Yes
2		2	No
3	Diabetes mellitus	1	Yes
3		2	No
4	Progesterone use	1	Yes
4		2	No
5	Spontaneous abortion	1	Yes
5		2	No
6	Conception method	1	Natural conception
6		2	IVF
7	Upper respiratory infection	1	Yes
7		2	No
8	NSAID use	1	Yes
8		2	No
9	Anticonvulsant use	1	Yes
9		2	No
10	Lithium use	1	Yes
10		2	No
11	ACEI use	1	Yes
11		2	No
12	Anemia	1	Yes
12		2	No
13	Pregnant woman smoking	1	Yes
13		2	No
14	Pregnant woman drinking	1	Yes
14		2	No
15	Spouse smoking	1	Yes
15		2	No
16	Spouse drinking	1	Yes

(Continued)

Table 1 (continued).

N of Variables	Variables	Value	Group details
16		2	No
17	Spouse's medication use during preparation for pregnancy	1	Yes
17		2	No
18	Exposure to radioactive substances	1	Yes
18		2	No
19	Mental stress	1	Yes
19		2	No
20	Intermarriage	1	Yes
20		2	No
21	Family history of CHD	1	Yes
21		2	No
22	Twin pregnancy	1	Single
22		2	Twin
23	Fetal arrhythmia	1	Yes
23		2	No
24	Fetal hydrops	1	Yes
24		2	No
25	Thyroid disorder	1	Yes
25		2	No
26	Gestational weeks	1	$16 \leq GW < 28$
26		2	$28 \leq GW < 40$

Fetal echocardiography was performed by experienced associate chief physicians and chief physicians, and diagnoses were then made based on grayscale and color images and pulse wave Doppler according to multiple section screening, including four-chamber, left and right ventricular outflow tract (LVOT and RVOT), three-vessel (3V), and three vessels and trachea (3VT) views as well as sagittal views of the superior and inferior vena cava, aortic arch, and ductal arch. All relevant physicians at the participating institutions were trained according to these guidelines. All images that were uploaded to our database from participating institutions were independently reviewed by two experienced physicians who confirmed or corrected the diagnoses.

2.4 Statistical Analysis

Continuous variables with a Gaussian distribution were expressed as the mean ± standard deviation and were compared using a t-test between the two groups, while non-Gaussian variables were expressed as the median (interquartile range, IQR) and were compared using a non-parameter test between the groups. Noncontinuous variables are expressed as percentages (%). Chi-square tests or Fisher's exact tests were performed to compare the variables between the two groups.

BN was performed to detect potential relationships between factors and fetal CHD, which consisted of two components: a directed acyclic graph (DAG) that encoded the dependency structure of the network and a

conditional probability table (CPT) for each node given its parent set. The learning process of the BN consisted of two parts: structural learning and parameter estimation. A BN can reconstruct a join probability distribution between variables. Let $X = (x_1, x_2, \ldots, x_n)$ represent n random variables. A DAG $G = (V, E)$ contains node set V and edge set E of the network, where node v_I represents a random variable x_I, and a directed edge $e_{ij} \in E$ from v_I to v_j indicates the probabilistic dependency relationship between node i and node j. Each node v_I holds a CPT that contains an entry for each joint assignment given a specific value of the parent set $Pa(x_I)$ of node v_i. Based on the independent assumption, each node x_I is independent of its nondescendants given its parent set $Pa(x_I)$ in the DAG. Therefore, a joint probability distribution of variable X can be decomposed as follows (Eq. (1)):

$$\left(P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Pa(x_i)) \right) \tag{1}$$

2.4.1 Build Bayesian Network Models

We construct a BN using the Matlab software (https://matlab.mathworks.com) in two steps. Firstly, learn the structure of the BN, i.e., identifying the topological structure of the BN; Secondly, learn the parameters of the BN, i.e., by estimating the CPT of each node from the dataset after the structure of the BN is identified. In order to simplify the model, BN is based on discrete variables here. The values of most of the exposure variables are discrete. The variables with more values, such as the gestational weeks and the age of the pregnant woman, are classified into several discrete variables with value to facilitate the learning of BN.

We applied four algorithms to learn the structure of the BN, which is now available via the open -source software BNT Murphy (2004) [12]. These algorithms are K2+T (K2 with two random initializations) with MWST (maximum weight-spanning tree) initialization), K2-T (K2 with MWST inverse initialization) GS+T (GS (starting from an empty structure) starting from a MWST-initialized structure) and GES (greedy search in the space of equivalent classes), each of which deals with complete data [13–15].

When the network topology G is confirmed by these four algorithms, the parameters of CPT at each node are learned by the maximum likelihood estimation (MLE) given the complete dataset.

2.4.2 Choosing a Best BN Model

We chose the best model that had a better representation of the original data distribution. K-fold cross-validation was applied to evaluate the performance of the model, in which the dataset was divided into K independent subsets, and every time the K-1 subset was applied to train the BN model, another subset was used to test the model that was just trained; after K times, ROC curves were drawn based on the prediction score of the classifying class (Figs. 1 and 2). The model with the highest average area under the ROC curve (AUC) for K times was chosen as the final BN model. Considering the prediction accuracy and the running time of the algorithm, we choose the K2-T algorithm to learn the structure of the BN, whose AUC was 0.594, higher than other algorithms (AUC for K2+T, GS+T and GES: 0.579, 0.588 and 0.574, with *P* value of 0.011, 0.314, and 0.002 respectively when compared with AUC for K2-T). We did find that GWs greatly improved the accuracy of prediction. Therefore, it is necessary to consider the probabilities at which different results can be obtained in in different GW group.

2.4.3 Probabilistic Reasoning Based on BN

When BN is established, causal reasoning can be conducted by the junction tree algorithm [16]. In cases where one or more nodes were observed, we could acquire the marginal probability distributions of the result node and the observed node and then utilize the Bayesian rule to obtain the probability of a specific value being obtained at the result node after providing evidence from the observed node.
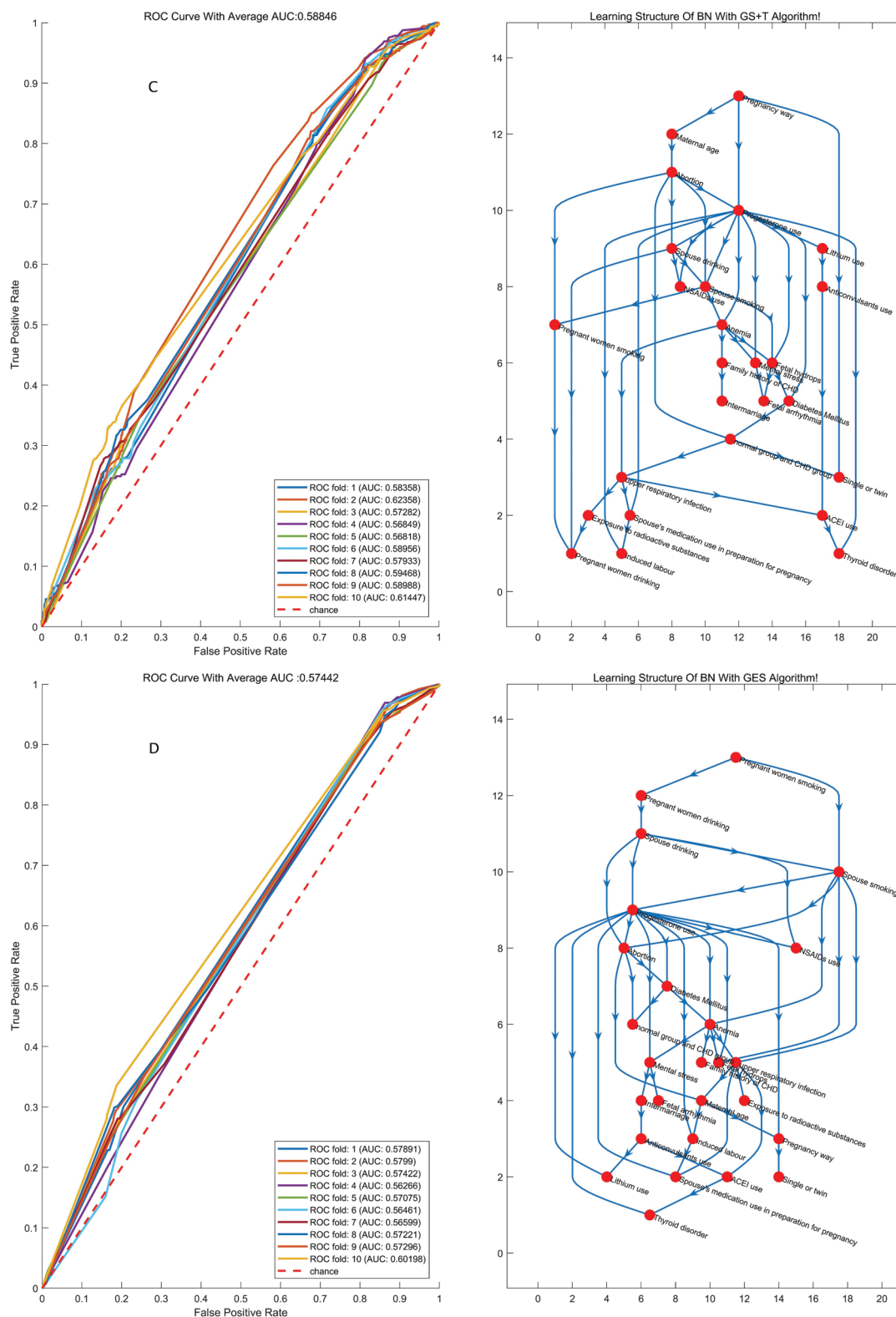
Figure 1: (continued)

Figure 1: The structures of the BN and the accuracy of the predictions of the four algorithms regardless of GW. BN: Bayesian network; MLE: Maximum likelihood estimation; AUC: Area under the curve
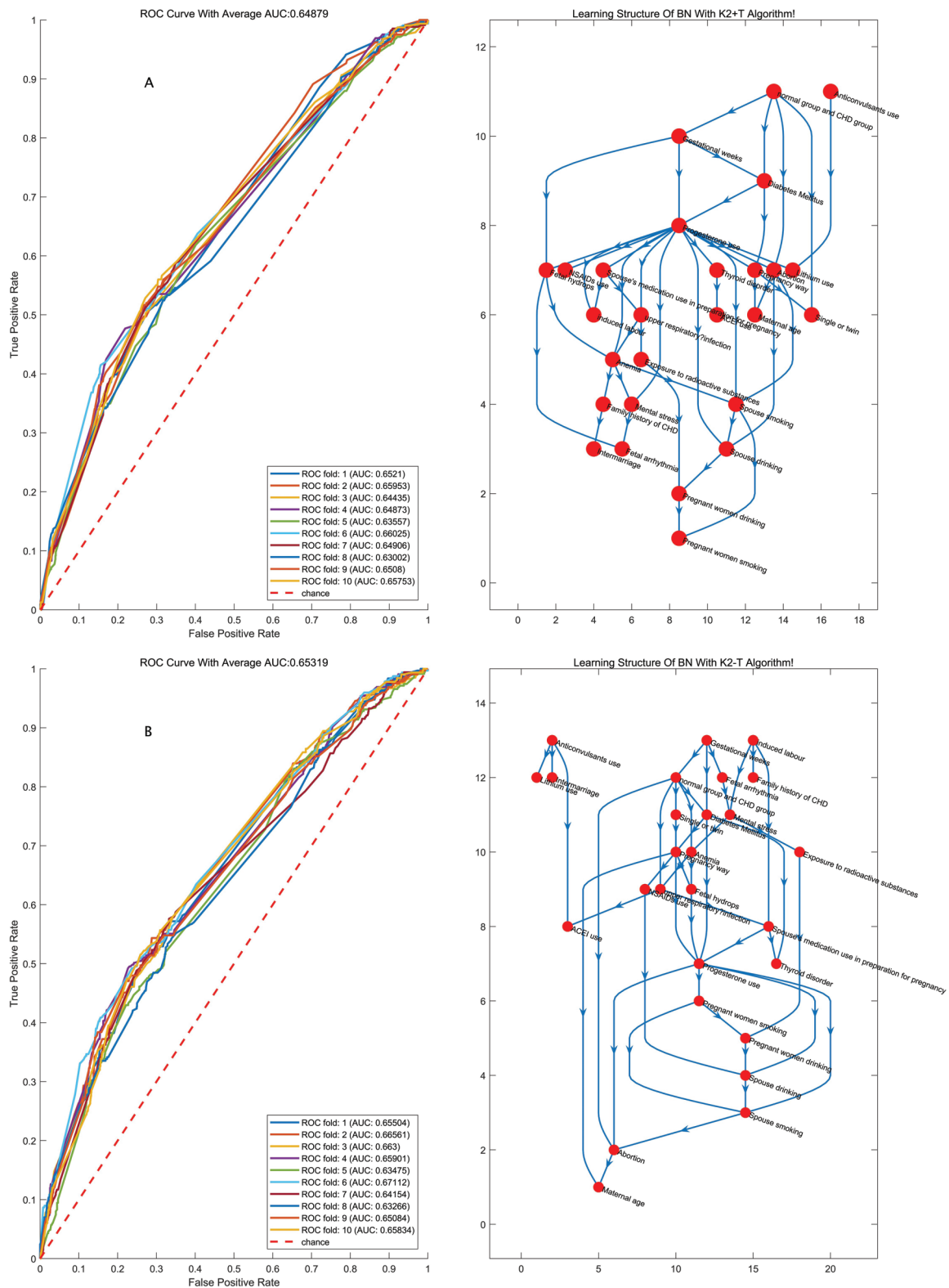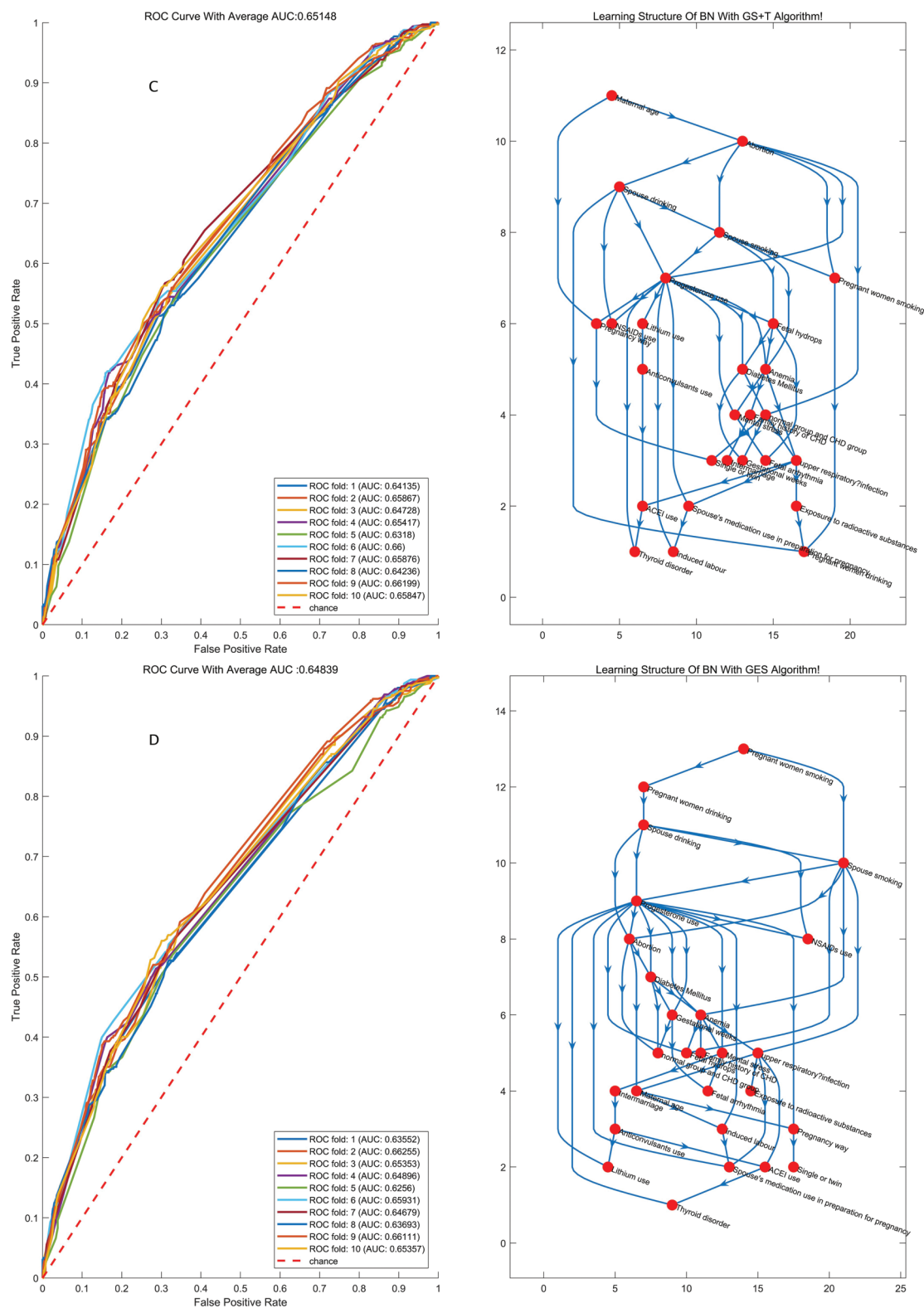
Figure 2: (continued)

Figure 2: The structures of the BN and the accuracy of the predictions of the four algorithms considering GW

We conducted two experiments. When GWs were ignored, the exposure group was set up, and we calculated the probability of the disease occurring given different combinations of observed factors based on causal reasoning under the condition of a node with no considering set to normal value. The probability that all nodes would be set to normal values as a result of the control group was determined, and the risk ratio (RR) could then be calculated by the predictive incidence probability in the exposure and control groups. When GWs were considered, the subjects were divided into two groups (Group A: $16 \leq, < 28$ weeks, and Group B: $28 \leq, < 40$ weeks).

To study whether there is a risk of disease caused by hazardous environmental factors, the experimental group and the control group are usually set up, and the risk ratio (Risk Ratio, RR) of the two groups is calculated, which represents the ratio of the incidence of the experimental group and the control group. The calculation process is shown in Eq. (2). In this formula, m_{11} and m_{12} represents the number of cases in the exposure group and in the control group, m_{21} and m_{22} representing the number of non-cases in both groups.

$$\left(RR = \frac{\dfrac{m_{11}}{(m_{11} + m_{21})}}{\dfrac{m_{12}}{(m_{12} + m_{22})}} = \frac{m_{11}m_{12} + m_{11}m_{22}}{m_{12}m_{11} + m_{12}m_{21}} \right) \tag{2}$$

In this article, the experimental group and the control group are distinguished by the values of environmental variables. In the Bayesian network model, the calculation of the risk ratio RR value is shown in Eq. (3).

$$\left(RR = \frac{P(y = 1 | x_1, x_2, \ldots, x_n)}{P(y = 1 | x_1', x_2', \ldots, x_n')} \right) \tag{3}$$

3 Results

3.1 Baseline Characteristics in the Normal Group and the Fetal CHD Group

Among all the participants, 3,312 pregnant women had CHD fetuses, and the remainders were normal fetuses. The age range of most pregnant women (approximately 80%) was between 20 and 35 years old. The proportion of subjects in group A (GWs < 28 weeks) was 75.9%, and the proportion in group B (GW $\geq$ 28 weeks) was 24.1%. The percentages of patients with diabetes, anemia, upper respiratory infection, progesterone use and family history of CHD were higher in the normal group than in the CHD group, while the rates of spontaneous abortion, spouse smoking and drinking, and twin pregnancy were higher in the CHD group (all $P < 0.05$, Tab. 2). All participants have a good control of the glucose level. Tab. 3 shows the detailed types of fetal CHD included in our study.

Table 2: Baseline characteristics of the normal group and the fetal CHD group

Group	Fetal CHD group	Normal group	P-value
N	3312	12774	–
Age (n, %)			0.048
<20 years old	18 (0.54%)	38 (0.30%)	
20≤,<35 years old	2746 (82.91%)	10728 (84.00%)	
≥35 years old	548 (16.55%)	2008 (15.72%)	

(Continued)

Table 2 (continued).

Group	Fetal CHD group	Normal group	P-value
Induced labor (n, %)			0.309
Yes	124 (3.74%)	432 (3.38%)	
No	3188 (96.26%)	12342 (96.62%)	
Diabetes mellitus			1.09×10^{-45}
Yes	145 (4.38%)	1680 (13.15%)	
No	3167 (95.62%)	11094 (86.85%)	
Gestational weeks			2.27×10^{-106}
16≤, <28	2034 (61.41%)	10178 (79.68%)	
28≤, <40	1278 (38.59%)	2596 (20.32%)	
Progesterone use			9.18×10^{-4}
Yes	542 (16.36%)	2410 (18.87%)	
No	2770 (83.64%)	10364 (81.13%)	
Spontaneous abortion			8.49×10^{-14}
Yes	918 (27.72%)	2760 (21.61%)	
No	2394 (72.28%)	10014 (78.39%)	
Conception method			0.694
Natural conception	3216 (97.10%)	12387 (96.97%)	
IVF	96 (2.90%)	387 (3.03%)	
Upper respiratory infection			0.016
Yes	641 (19.35%)	2717 (21.27%)	
No	2671 (80.65%)	10057 (78.73%)	
NSAID use			0.511
Yes	22 (0.66%)	99 (0.78%)	
No	3290 (99.34%)	12675 (99.22%)	
Anticonvulsant use			0.611
Yes	0 (0.00%)	1 (0.01%)	
No	3312 (100.00%)	12773 (99.99%)	
Lithium use			0.471
Yes	0 (0.00%)	2 (0.02%)	
No	3312 (100.00%)	12772 (99.98%)	
ACEI			0.145
Yes	3 (0.09%)	4 (0.03%)	
No	3309 (99.91%)	12770 (99.97%)	
Anemia			8.30×10^{-6}
Yes	242 (7.31%)	1256 (9.83%)	

(Continued)

Table 2 (continued).

Group	Fetal CHD group	Normal group	P-value
No	3070 (92.69%)	11518 (90.17%)	
Pregnant woman smoking			0.080
Yes	9 (0.27%)	64 (0.50%)	
No	3303 (99.73%)	12710 (99.50%)	
Pregnant woman drinking			0.579
Yes	11 (0.33%)	51 (0.40%)	
No	3301 (99.67%)	12723 (99.60%)	
Spouse smoking			9.26×10^{-5}
Yes	428 (12.93%)	1379 (10.8%)	
No	2884 (87.08%)	11395 (89.20%)	
Spouse drinking			5.50×10^{-4}
Yes	218 (6.58%)	654 (5.12%)	
No	3094 (93.42%)	12120 (94.88%)	
Spouse's medication use during preparation for pregnancy			0.865
Yes	81 (2.45%)	319 (2.50%)	
No	3231 (97.55%)	12455 (97.50%)	
Exposure to radioactive substances			0.115
Yes	26 (0.79%)	140 (1.10%)	
No	3286 (99.21%)	12634 (98.90%)	
Mental stress			0.271
Yes	79 (2.39%)	265 (2.07%)	
No	3233 (97.61%)	12509 (97.93%)	
Intermarriage			0.283
Yes	2 (0.06%)	3 (0.02%)	
No	3310 (99.94%)	12771 (99.98%)	
Family history of CHD			0.030
Yes	24 (0.72%)	148 (1.16%)	
No	3288 (99.28%)	12626 (98.84%)	
Twin pregnancy			5.72×10^{-5}
Single	3235 (97.68%)	12601 (98.65%)	
Twins	77 (2.32%)	173 (1.35%)	
Fetal arrhythmia			0.950
Yes	40 (1.21%)	156 (1.22%)	
No	3272 (98.79%)	12618 (98.78%)	

(Continued)

Table 2 (continued).

Group	Fetal CHD group	Normal group	*P*-value
Fetal hydrops			0.124
Yes	72 (2.17%)	338 (2.65%)	
No	3240 (97.83%)	12436 (97.35%)	
Thyroid disorder			0.467
Yes	19 (0.57%)	88 (0.69%)	
No	3293 (99.43%)	12686 (99.31%)	

Table 3: The types of fetal CHD observed in this study

Type of fetal CHD	N (%)
Congenital anomaly of systemic, pulmonary, or umbilical veins or ductus venosus	323 (9.75)
Congenital anomaly of an atrium or atrial septum	89 (2.69)
Congenital anomaly of an atrioventricular connection	174 (5.25)
Congenital anomaly of a ventricle or the ventricular septum	632 (19.08)
Congenital anomaly of a ventriculo-arterial connection	784 (23.67)
Congenital anomaly of the great arteries, including the arterial duct	598 (18.06)
Fetal cardiomyopathy	11 (0.33)
Fetal cardiac tumors	86 (2.60)
Fetal arrhythmia	222 (6.70)
Others	393 (11.87)
Total	3312 (100)

Others included ventricular outpouchings, aneurysm of the atrial appendage and atrium, abnormal heart position, abnormal cardiothoracic or atrial-ventricular ratio, abnormal proportions of the aorta and pulmonary artery, heart failure, pericardial tumors, multiple intracardiac echogenic focus or multiple calcifications, bicuspid aortic valves.

3.2 Effect Analysis of Associated Factors Ranging from Single- to Multiple-Factor Exposure Based on the BN

The BN structure analysis showed that several factors were directly associated with fetal CHD; these factors included a history of spontaneous abortion, upper respiratory tract infection during early pregnancy, anemia, and mental stress as well as twin pregnancy and parental smoking. Based on the causal reasoning of the BN, we found that the risk of fetal CHD gradually increased with potentially synergistic exposure of ranging from a single factor to multiple factors. A single factor analysis demonstrated that the RRs of twin pregnancy, spontaneous abortion, or spouse smoking were 1.50, 1.38, and 1.11 (all $P < 0.001$), respectively. When twin pregnancy was combined with spontaneous abortion or spouse smoking, the RR for fetal CHD was greater (RR = 1.96 or 1.64, $P < 0.05$) than that of single factor exposure. The risk continued to increase when three factors, such as a combination including anemia and upper respiratory infection or the addition of twin pregnancy or spontaneous abortion or spouse smoking, were superimposed, resulting in RRs of 1.56, 1.44 or 1.17 (all $P < 0.05$), respectively.

By this analogy, when we combined four factors, including twin pregnancy, upper respiratory tract infection during early pregnancy, anemia, and mental stress as well as spontaneous abortion or spouse smoking, the risk increased ranging from 1.67 to 2.12 ($P < 0.05$). When the above factors were combined into a set of five factors, the risk was increased to up to 1.62 times (RR = 2.62 or 2.28, $P < 0.001$), higher than the risk for subjects with less than 5 factors (Tab. 4). Aside from the factors mentioned above, other factors did not continue to increase the risk of fetal CHD.

Table 4: Risk ratio of the combined exposure variables for the fetal CHDs in different groups

Exposure factor group	Subject group	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Predictive CHD incidence probability in CHD group	Predictive CHD incidence probability in control group	Risk ratio (RR)	*P*-value
Single factor exposure	Total population	Single or twin = 2	/	/	/	/	0.3244	0.2169	1.496	5.72×10^{-5}
		Abortion = 1	/	/	/	/	0.2999	0.2169	1.383	8.49×10^{-14}
		Spouse smoking = 1	/	/	/	/	0.2415	0.2169	1.113	5.50×10^{-4}
	$16 \leq$ gw < 28	Single or twin = 2	/	/	/	/	0.2675	0.1740	1.537	0.162
		Abortion = 1	/	/	/	/	0.2458	0.1740	1.412	0.879
		Spouse smoking = 1	/	/	/	/	0.1949	0.1740	1.120	0.101
	$28 \leq$ gw < 40	Single or twin = 2	/	/	/	/	0.4959	0.3620	1.370	3.54×10^{-9}
		Abortion = 1	/	/	/	/	0.4674	0.3620	1.291	9.55×10^{-43}
		Spouse smoking = 1	/	/	/	/	0.3947	0.3620	1.090	2.25×10^{-21}
Synergistic effect of two factors	Total population	Abortion = 1	Single or twin = 2	/	/	/	0.4262	0.2169	1.965	0.039
		Spouse smoking = 1	Single or twin = 2	/	/	/	0.3556	0.2169	1.640	0.025
	$16 \leq$ gw < 28	Abortion = 1	Single or twin = 2	/	/	/	0.3610	0.1740	2.074	0.198
		Spouse smoking = 1	Single or twin = 2	/	/	/	0.2956	0.1740	1.699	0.764
	$28 \leq$ gw < 40	Abortion = 1	Single or twin = 2	/	/	/	0.6034	0.3620	1.667	0.041
		Spouse smoking = 1	Single or twin = 2	/	/	/	0.5307	0.3620	1.466	0.585
Synergistic effect of three factors	Total population	upper respiratory infection = 1	Anemia = 1	Single or twin = 2	/	/	0.3386	0.2169	1.561	0.029
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	/	/	0.3136	0.2169	1.446	0.023
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	/	/	0.2534	0.2169	1.168	0.013
	$16 \leq$ gw < 28	upper respiratory infection = 1	Anemia = 1	Single or twin = 2	/	/	0.2803	0.1740	1.611	0.146

(Continued)

Table 4 (continued).

Exposure factor group	Subject group	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Predictive CHD incidence probability in CHD group	Predictive CHD incidence probability in control group	Risk ratio (RR)	P-value
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	/	/	0.2578	0.1740	1.482	0.938
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	/	/	0.2052	0.1740	1.179	0.236
	$28 \leq gw < 40$	upper respiratory infection = 1	Anemia = 1	Single or twin = 2	/	/	0.5120	0.3620	1.414	0.049
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	/	/	0.4834	0.3620	1.335	4.16×10^{-5}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	/	/	0.4102	0.3620	1.133	7.85×10^{-6}
Synergistic effect of four factors	Total population	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	Single or twin = 2	/	0.4607	0.2169	2.124	2.59×10^{-7}
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	/	0.4325	0.2169	1.994	4.14×10^{-6}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Single or twin = 2	/	0.3704	0.2169	1.708	7.51×10^{-4}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Mental stress = 1	/	0.3615	0.2169	1.667	0.001
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	Single or twin = 2	/	0.4419	0.2169	2.037	1.69×10^{-6}
	$16 \leq gw < 28$	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	Single or twin = 2	/	0.3938	0.1740	2.263	1.09×10^{-6}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Single or twin = 2	/	0.3092	0.1740	1.777	0.002
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Mental stress = 1	/	0.3010	0.1740	1.730	0.003
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	Single or twin = 2	/	0.3759	0.1740	2.160	6.13×10^{-6}
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	/	0.3669	0.1740	2.109	1.41×10^{-5}
	$28 \leq gw < 40$	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	Single or twin = 2	/	0.6364	0.3620	1.758	4.06×10^{-8}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Single or twin = 2	/	0.5466	0.3620	1.510	2.09×10^{-4}

(Continued)

Table 4 (continued).

Exposure factor group	Subject group	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Predictive CHD incidence probability in CHD group	Predictive CHD incidence probability in control group	Risk ratio (RR)	P-value
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Mental stress = 1	/	0.5371	0.3620	1.483	4.32×10^{-4}
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	Single or twin = 2	/	0.6187	0.3620	1.709	2.82×10^{-7}
		Abortion = 1	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	/	0.6096	0.3620	1.684	7.27×10^{-7}
Synergistic effect of five factors	Total population	Abortion	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	Single or twin = 2	0.5692	0.2169	2.624	5.48×10^{-13}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Mental stress = 1	Single or twin = 2	0.4954	0.2169	2.284	6.03×10^{-9}
	$16 \leq gw < 28$	Abortion = 1	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	Single or twin = 2	0.5012	0.1740	2.880	4.54×10^{-12}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Mental stress = 1	Single or twin = 2	0.4275	0.1740	2.457	3.24×10^{-8}
	$28 \leq gw < 40$	Abortion = 1	upper respiratory infection = 1	Anemia = 1	Mental stress = 1	Single or twin = 2	0.7302	0.3620	2.017	1.41×10^{-13}
		upper respiratory infection = 1	Anemia = 1	Spouse smoking = 1	Mental stress = 1	Single or twin = 2	0.6679	0.3620	1.845	9.31×10^{-10}

3.3 Effect Analysis of Associated Factors Ranging from Single- to Multiple-Factor Exposure Based on the BN When Grouped by GWs

We further performed a sensitivity analysis in the participants grouped by GWs. We obtained a result consistent with the above-described result no matter we explored the population in Group A or Group B or used a single factor or multifactor exposure analysis. These findings demonstrate that the risk of a combination of these five factors occurring was as high as 2.88 ($P < 0.05$) in Group A and as high as 2.02 ($P < 0.05$) in Group B when the involved factors included twin pregnancy, upper respiratory tract infection during early pregnancy, anemia, and mental stress as well as a history of spontaneous abortion. The detailed risk ratios for these factors are presented in Tab. 4.

4 Discussions

4.1 The Main Findings of the Study

This study focused on factors affecting fetal CHD in a large sample size consisting of 3,312 pregnant women with CHD fetuses among 16,086 subjects. Importantly, instead of traditional logistic regression analysis, we used BN analysis, a method based on probabilistic reasoning, to explore the interactions among specific factors and factors associated with an increased risk of fetal CHD. The results of this research have revealed that the factors directly associated with fetal CHD included history of spontaneous

abortion, upper respiratory tract infection during early pregnancy, anemia, and mental stress as well as twin pregnancy and parental smoking. The synergistic exposure of more factors increased the risk of fetal CHD, leading to RR as high as 2.62 for the five-factor synergistic effect. In addition, a sensitivity analysis of groups divided by GWs demonstrated the risk showed the same trends as those described above from the single factor to synergistically multiple-factor exposure in both groups.

4.2 Progress and Differences in the Analysis of Risk Factors for Fetal CHD

The factors associated with fetal CHD identified in this study have also been described in other similar studies. One of these factors is maternal illness, such as upper respiratory tract infection during early pregnancy. A meta-analysis of maternal viral infection and the risk of fetal CHD suggested that mothers who had history of viral infection in early pregnancy had significantly higher risk of having offspring with CHD (RR = 2.28), and this risk was more significant in mothers with rubella and cytomegalovirus. The effect of nonspecific maternal infection is difficult to definitively separate from the effects of medications used to treat the illness, including maternal fever and infection [17]. Jenkins et al. [18] reported an up to 1.9-fold increase in the risk of aggregate cardiac defects in patients with maternal febrile illness and a 1.1-fold increase in the rate of any heart defects among subjects with maternal influenza infection in early pregnancy. The results of these studies are consistent with ours. However, the viruses were not classified in detail in our study.

In addition to maternal infection, other maternal chronic diseases associated with a risk of CHD in offspring, including diabetes, hypertension, anemia, connective tissue disorders, epilepsy and mood disorders, have been reported to be significantly associated with a higher prevalence of any form of CHD in offspring. Moreover, the population-attributed risk for CHD was investigated, and the results showed that the highest population-attributable risk was noted for anemia (2.17%), followed by type 2 diabetes (1.45%) and hypertension (0.71%) [19]. Similar results were reported by Liu et al. [20]. Anemia is a complex condition that may be associated with many factors, which should be interpreted carefully to determine the potential causes for the risk of CHD and the possibility that confounding factors, such as multivitamin and folate supplementation or malnutrition, may affect results. The other environmental risk factor associated with CHD was maternal mental stress (odds ratios ranged from 2.48 to 3.93) during early pregnancy, which has also been reported in two previous hospital-based case-control studies [21,22]. They are similar to the results of our study. Although the biological mechanisms by which maternal stress causes CHD are not clear, we strongly suggest that psychological management for pregnant women be strengthened, especially during early pregnancy. Spontaneous abortion was found to be associated with the risk of CHD in our study. The causal relationship could not be inferred according to current knowledge, and these results only indicate that a history of miscarriage is a predictor of having an infant born with CHD or an increased risk of tetralogy of Fallot [22,23]. In any case, these findings suggest that the management of obstetric healthcare and counseling for women with a history of miscarriages should be strengthened to reduce the incidence of CHD.

The association between maternal diabetes and an increased risk of CHD has been clearly described in many studies [7,19,20]. However, no similar finding was obtained in our research. This may be because of selection bias in the population recruited in our center. Many pregnant women with diabetes who have a good control of blood glucose level are referred to our center from local hospitals for further fetal echocardiography examination, and most of these fetuses are normal, resulting in a nonrandomly selected population. This could further affect our conclusion.

The correlation between paternal smoking and congenital cardiovascular defects has been studied, but too little information is available to determine the associated risk. Of the many congenital defects observed in a nursery, there was a significantly higher incidence of cardiovascular system abnormalities in the tobacco-exposed group [24]. A case-control study suggested that there is an association between periconceptional

tobacco exposure and an increased risk of CHD during the neonatal period and that there may be a dose effect [25]; however, this needs to be confirmed in a larger population. Unfortunately, in our study, we were not able to verify this dose-effect relationship, although our results do suggest that paternal smoking is a risk factor for CHD. The potential mechanisms underlying the teratogenicity associated with periconceptional tobacco exposure remain unclear. One possible reason is that nicotine and carbon monoxide damage placental functions, leading to fetal hypoxia [26,27].

With regard for fetal factors, we found that there was a correlation between multiple gestations and fetal CHD. In 2016, Panagiotopoulou et al. conducted a study on CHD in twin pregnancy and showed that monochorionicity (OR 3.49, 95% CI 1.57–7.77) was a significant determinant of CHD that was independent of maternal age, parity, and the gender of the offspring [28]. In another study, when compared with singletons, twins were at increased risk of CHD, and the risk was substantially higher among monochorionic twins. These findings are important for doctors counselling women with twin pregnancies [29].

4.3 Advantages and Limitations of This Study

The large sample size included in this study makes it the first study to assess factors related to fetal CHD using BN analysis, which can visually reflect potential relationships among multiple factors by constructing a DAG. There is no strict requirement for statistical assumptions. However, conclusions from previous studies have usually been drawn based on logistic regression models and the premise that the evaluated factors are independent with each other. Based on the current results, the interactions among factors and the synergistic effects of multiple factors on the risk of CHD are difficult to consider and accurately estimate. Therefore, considering the large sample dataset and the interactions observed between the factors included in this study, a BN is a very reasonable method for exploring factors that affect CHD, and these results are objective and in agreement with medical explanations. This method not only selected several factors that were clearly related to fetal CHD but also presented an accurate estimation of factors and the synergistic effect of multiple factors to compensate for the shortcomings of current research. Another strength of the study is the fact that we examine fetal CHD as opposed to only live births and thus would capture pregnancies who would go on to have intrauterine demise or termination that would not be captured in a neonatal/live birth registry. We acknowledge that although it based on a large population, the data were mainly obtained from self-reported questionnaires, and the accuracy of information collection is therefore a problem that needs to be considered. Moreover, this is a cross-sectional study that demonstrates only the correlations between these factors and fetal CHD but does not provide causal relationships. One additional limitation is that our center is a referral center for fetal heart disease. The fetuses referred to our center come from all over the country. Therefore, some of the patients coming to our center are pregnant women with known risk factors or with a fetus previously found to have CHD at a local hospital, and this may have led to selection bias in the population. In addition, fetal CHD was diagnosed by fetal echocardiography and we did not make postnatal verification for every case. But our findings can be credible, because fetal echocardiographic diagnoses were mostly consistent with autopsy findings in our center [30], which has shown that fetal cardiovascular anomalies disclosed by FE were completely in line with autopsy findings in 87.1% (149/171) of cases. A final limitation may be that the relationships between genetic factors and fetal CHD were not considered in this study, which is solely focused on clinical characteristics. Considering the factors above, our findings should be interpreted cautiously and may not be generalizable to the general population.

5 Conclusions

The structured learning and parameter estimation in the BN demonstrated that factors directly associated with CHD included a history of spontaneous abortion, upper respiratory tract infection during early

pregnancy, anemia, and mental stress as well as the number of births and spouse smoking. Combinations containing higher numbers of these factors were associated with a higher risk of fetal CHD for the total population or the population grouped by GWs. All these findings suggest that improvements in the management of obstetric healthcare and the provision of prenatal counseling for women with these risk factors should be strengthened to decrease the incidence of CHD.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. GBD 2017 Congenital Heart Disease Collaborators (2017). Global, regional, and national burden of congenital heart disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Child & Adolescent Health, 4(3),* 185–200, 2020.

2. Lin, K. Y., D'Alessandro, L. C., Goldmuntz, E. (2013). Genetic testing in congenital heart disease: Ethical considerations. *World Journal for Pediatric & Congenital Heart Surgery, 4(1),* 53–57. DOI 10.1177/2150135112459523.

3. Tennant, P. W., Pearce, M. S., Bythell, M., Rankin, J. (2010). 20-year survival of children born with congenital anomalies: A population-based study. *Lancet, 375(9715),* 649–656. DOI 10.1016/S0140-6736(09)61922-X.

4. Wren, C., Irving, C. A., Griffiths, J. A., O'Sullivan, J. J., Chaudhari, M. P. et al. (2012). Mortality in infants with cardiovascular malformations. *European Journal of Pediatrics, 171(2),* 281–287. DOI 10.1007/s00431-011-1525-3.

5. Moons, P., Sluysmans, T., De Wolf, D., Massin, M., Suys, B. et al. (2009). Congenital heart disease in 111 225 births in Belgium: Birth prevalence, treatment and survival in the 21st century. *Acta Paediatrica, 98(3),* 472–477. DOI 10.1111/j.1651-2227.2008.01152.x.

6. Bruneau, B. G. (2008). The developmental genetics of congenital heart disease. *Nature, 451(7181),* 943–948. DOI 10.1038/nature06801.

7. Donofrio, M. T., Moon-Grady, A. J., Hornberger, L. K., Copel, J. A., Sklansky, M. S. et al. (2014). Diagnosis and treatment of fetal cardiac disease: A scientific statement from the American Heart Association. *Circulation, 129(21),* 2183–2242. DOI 10.1161/01.cir.0000437597.44550.5d.

8. García-Herrero, S., Mariscal, M. A., Gutiérrez, J. M., Ritzel, D. O. (2013). Using Bayesian networks to analyze occupational stress caused by work demands: Preventing stress through social support. *Accident Analysis and Prevention, 57(3),* 114–123. DOI 10.1016/j.aap.2013.04.009.

9. American Diabetes Association (2000). Gestational diabetes mellitus. *Diabetes Care, 23,* S77–S79.

10. American College of Obstetricians and Gynecologists (2008). ACOG Practice Bulletin No. 95: Anemia in pregnancy. *Obstetrics and Gynecology, 112(95),* 201–207.

11. Carvalho, J. S., Allan, L. D., Chaoui, R., Copel, J. A., DeVore, G. R. et al. (2013). ISUOG Practice Guidelines (updated): Sonographic screening examination of the fetal heart. *Ultrasound in Obstetrics & Gynecology, 41(3),* 348–359.

12. Leray, P., Francois, O. (2004). *BNT structure learning package: Documentation and experiments.* Laboratoire PSI, Universitè et INSA de Rouen. Tech. Rep.

13. Cooper, G. F., Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning, 9(4),* 309–347.

14. Heckerman, D., Geiger, D., Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning, 20(3),* 197–243.

15. Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research, 3,* 507–554.

16. Huang, C., Darwiche, A. (1996). Inference in belief networks: A procedural guide. *International Journal of Approximate Reasoning, 15(3),* 225–263. DOI 10.1016/S0888-613X(96)00069-2.

17. Ye, Z., Wang, L., Yang, T., Chen, L., Wang, T. et al. (2019). Maternal viral infection and risk of fetal congenital heart diseases: A meta-analysis of observational studies. *Journal of the American Heart Association, 8(9),* e011264. DOI 10.1161/JAHA.118.011264.

18. Jenkins, K. J., Correa, A., Feinstein, J. A., Botto, L., Britt, A. E. et al. (2007). Noninherited risk factors and congenital cardiovascular defects: Current knowledge: A scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: Endorsed by the American Academy of Pediatrics. *Circulation, 115(23),* 2995–3014. DOI 10.1161/CIRCULATIONAHA.106.183216.

19. Chou, H. H., Chiou, M. J., Liang, F. W., Chen, L. H., Lu, T. H. et al. (2016). Association of maternal chronic disease with risk of congenital heart disease in offspring. *Canadian Medical Association Journal, 188(17–18),* E438–E446. DOI 10.1503/cmaj.160061.

20. Liu, S., Joseph, K. S., Lisonkova, S., Rouleau, J., van den Hof, M. et al. (2013). Association between maternal chronic conditions and congenital heart defects: A population-based cohort study. *Circulation, 128(6),* 583–589. DOI 10.1161/CIRCULATIONAHA.112.001054.

21. Li, H., Luo, M., Zheng, J., Luo, J., Zeng, R. et al. (2017). An artificial neural network prediction model of congenital heart disease based on risk factors: A hospital-based case-control study. *Medicine, 96(6),* e6090. DOI 10.1097/MD.0000000000006090.

22. Liu, S., Liu, J., Tang, J., Ji, J., Chen, J. et al. (2009). Environmental risk factors for congenital heart disease in the Shandong Peninsula, China: A hospital-based case-control study. *Journal of Epidemiology, 19(3),* 122–130. DOI 10.2188/jea.JE20080039.

23. Tikkanen, J., Heinonen, O. P. (1992). Congenital heart disease in the offspring and maternal habits and home exposures during pregnancy. *Teratology, 46(5),* 447–454. DOI 10.1002/tera.1420460509.

24. Woods, S. E., Raju, U. (2001). Maternal smoking and the risk of congenital birth defects: A cohort study. *Journal of the American Board of Family Practice, 14(5),* 330–334.

25. Karatza, A. A., Giannakopoulos, I., Dassios, T. G., Belavgenis, G., Mantagos, S. P. et al. (2011). Periconceptional tobacco smoking and isolated congenital heart defects in the neonatal period. *International Journal of Cardiology, 148(3),* 295–299. DOI 10.1016/j.ijcard.2009.11.008.

26. Rogers, J. M. (2009). Tobacco and pregnancy. *Reproductive Toxicology, 28(2),* 152–160. DOI 10.1016/j.reprotox.2009.03.012.

27. Shi, M., Wehby, G. L., Murray, J. C. (2008). Review on genetic variants and maternal smoking in the etiology of oral clefts and other birth defects. *Birth Defects Research. Part C, Embryo Today: Reviews, 84(1),* 16–29. DOI 10.1002/bdrc.20117.

28. Panagiotopoulou, O., Fouzas, S., Sinopidis, X., Mantagos, S. P., Dimitriou, G. et al. (2016). Congenital heart disease in twins: The contribution of type of conception and chorionicity. *International Journal of Cardiology, 218,* 144–149. DOI 10.1016/j.ijcard.2016.05.029.

29. Best, K. E., Rankin, J. (2015). Increased risk of congenital heart disease in twins in the North of England between 1998 and 2010. *Heart (British Cardiac Society), 101(22),* 1807–1812.

30. Gao, S., Han, J., Yu, S., Guo, Y., Ruan, Y. et al. (2019). Comparison of fetal echocardiogram with fetal cardiac autopsy findings in fetuses with congenital heart disease. *Journal of Maternal–Fetal & Neonatal Medicine, 6(1),* 1–7. DOI 10.1080/14767058.2019.1700498.