Effectiveness of Folate-Oriented Tertiary Interventions on Prevention of Birth Defects (FOID)

Statistical Analysis Plan

The Data Management and Statistical Analysis Plan is directed to support the aims of the FOID Study

Clinical trials.gov Number	NCT03725878
Protocol version and date	August 20, 2020
SAP version and date	Version 3.0, March 19, 2024

SAP version history		
Version Date	**SAP Version**	**Details of Changes**
January 02, 2019	Version 1.0	
August 19, 2022	Version 1.1	Description of introduction, outcomes, case definitions and interventions with more details.
August 20, 2022	Version 1.2	Definitions of analysis populations, analysis of primary outcome and count secondary outcomes, and handling of missing data.
August 22, 2022	Version 2.0	Revisions after comments by Prof Duolao Wang, list of summary tables
March 19, 2024	Version 3.0	Additions to detailed definitions of outcomes and analysis population, and methods for covariate analysis.

Signatures		
Name and role	**Signature**	**Date**
Prof Guoying Huang (Chief investigator)	Guoying Huang	2024/3/25
Prof Weili Yan (Chief investigator & Trial Statistician)	Weili Yan	2024/03/20
Prof Duolao Wang (Senior Statistician)	Duolao Wang	2024/03/19
Yalan Dou, Yuanchen He (Trial Statisticians)	Yalan Dou Yuanchen He	2024/03/19

TABLE OF CONTENTS

1. INTRODUCTION

Birth defects are the main cause of fetal death, infant mortality, and morbidity worldwide. The protective effect of sufficient folic acid supplementations on several birth defects is well-established [1-3]. However, most women failed to achieve the optimal folate status from diet alone[4]. We assume that achieving recommended folate levels before pregnancy through awareness of insufficient status and timely initiation of folic acid supplementation, based on the current red blood cell (RBC) folate status and genetic background of women of reproductive age, will exert an effect on primary prevention of both fetus and birth defects.

The FOID study is a single-blind, two-arm cluster-randomized controlled trial which conducted in Shanghai and Jiangsu Kunshan city, China[5]. This trial aims to evaluate the effectiveness of a comprehensive folate-oriented tertiary interventions (before pregnancy, during pregnancy and after delivery) embedded in the current routine primary care system in reducing the incidence of fetus and birth defects among a preparing-for-pregnancy population. The control arm will receive the current routine standard antenatal care interventions before pregnancy, while the experimental arm will additionally receive a folate-oriented tertiary intervention based on this. Before conception, the participants in the experimental arm will receive an RBC folate concentration test at enrollment, and those with insufficient levels (<400ng/mL) will receive a phone interview with an instruction of folic acid supplementation to improve their folate status to the target level before pregnancy based on a second RBC folate examination. During pregnancy, pregnancies with positive tests in the routine nuchal translucency (NT) screen at early gestation, ultrasound screening in mid-gestation will be referred to the authorized tertiary hospital for counseling according to their genetic and prenatal diagnosis of the fetal defect(s). After delivery, newborns with fetus and birth defect(s) will be referred to a clinical team at the authorized tertiary pediatric hospital for treatment and follow-up. The current pragmatic cluster randomized trial will provide real-world evidence on the necessity and effectiveness of preconception intervention focusing on achieving sufficient folate nutrition levels in prevention of

fetal and birth defects.

The purpose of this Statistical Analysis Plan is to define the outcome variables, statistical methods, and analysis strategies to address the objectives of the FOID study.

2. STUDY OBJECTIVES AND OUTCOMES

2.1. Study Objectives

2.1.1. Primary objective

The primary objective is to evaluate the effectiveness of community-based folate-oriented tertiary interventions focusing on achieving sufficient RBC folate levels before conception among pregnant-planning women on the incidence of fetus and birth defects.

2.1.2. Secondary objectives

The secondary objectives include to determine the effectiveness of the experimental interventions compared with the routine primary care on the following outcomes: (1) the incidence rate of fetus defects, birth defects, spontaneous pregnancy loss (including miscarriages and artificial abortions) and stillbirths related to all kinds of congenital defects; (2) the incidence of infant death or severe organ dysfunctions; and (3) prevalence of maternal gestational complications, including gestational diabetes, hypertension, and preeclampsia; (4) the medical costs related to fetus and birth defects during pregnancy and after birth.

2.2. Outcomes

2.2.1. Primary outcome

The primary outcome is the occurrence of fetus defects, stillbirth, and neonatal birth defects identified from the confirmation of pregnancy to 28 days after birth.

- Description:

  This is a composite outcome that includes a total of 24 types of defects according to the national birth defects surveillance policy, including fetus defects detected by

screening examinations for Down syndrome (DS), NT during early gestation, organ defects identified by routine antenatal ultrasound image examinations during the second trimester, stillbirth occurring during the entire gestation period, and birth defects identified after delivery and diagnosed by clinical examinations including ultrasound examination (Details of 24 defects types see Appendix 1). The primary outcome will be defined as "occurred" if any one of the 24 defects occurs. All examinations to measure the primary outcome will follow current routine protocols. Besides, detailed information on both miscarriages and artificial abortions will be collected, including the time of abortion, reasons for abortion, and antenatal information before an abortion.

- Time Frame: from the confirmation of pregnancy to 28 days after birth.
- Type: binary variable (Yes / No).

2.2.2. Secondary outcomes

(1) The occurrence of fetus defects

- This is a composite outcome and will be defined as "occurred" if any of the 24 defects are detected by routine ultrasound screening during pregnancy (Details of defects types see Appendix 1).
- Time Frame: from the confirmation of pregnancy to before delivery.
- Type: binary variable (Yes / No).

(2) The occurrence of birth defects

- This is a composite outcome and will be defined as "occurred" if any of the 24 defects are detected by ultrasound examination within one year old (Details of defects types see Appendix 1).
- Time Frame: from birth to one year old.
- Type: binary variable (Yes / No).

(3) The occurrence of infant death or severe organ dysfunctions

- This is a composite outcome and will be defined as “occurred” if either infant death or severe organ dysfunctions occur.
- Time Frame: from birth to 6 months after delivery (can be expanding to the end of the 7th month).
- Type: binary variable (Yes / No).

(4) The occurrence of abortion that related with congenital defects

- Time Frame: from the confirmation of pregnancy to the 28th gestational week.
- Type: binary variable (Yes / No).

(5) Count of fetus defects, stillbirth, and birth defects

- Time Frame: from the confirmation of pregnancy to 28 days after birth.
- Type: Count-Number of events.

(6) Occurrence of confirmed congenital heart defects (CHD)

- All pregnancies will have prenatal ultrasound scans in the second trimester (around 20 to 24 weeks of gestation) performed by trained ultra sonographers; abnormal findings will be confirmed by authorized tertiary fetal medical centers. All newborns will have routine neonatal CHD screening, and those screened positive will be subsequently sent for further echocardiographic confirmation. The nomenclature of the International Paediatric and Congenital Cardiac Code (of the Nomenclature Working Group of the International Society for Nomenclature of Paediatric and Congenital Heart Disease) will be used to define CHD.
- Time Frame: from the confirmation of pregnancy to 28 days after birth.
- Type: binary variable (Yes / No).

(7) Occurrence of maternal gestational diabetes (GDM)

- Pregnant participants will receive a 75-g oral glucose tolerance test during 24-28 gestational weeks, and GDM will be diagnosed if any of the following criteria are met, based on the recommendations of the International Association of the Diabetes and Pregnancy Study Groups Consensus Panel: fasting glucose ≥5.1 mmol/L, 1-hour glucose ≥10.0 mmol/L, or 2-hour glucose ≥8.5 mmol/L.
- Time Frame: during 24-28 gestational weeks.
- Type: binary variable (Yes / No).

(8) Maternal gestational weight gain (kg)

- This outcome will be calculated as body weight at delivery minus weight before pregnancy (self-reported).
- Time Frame: from pre pregnancy to delivery.
- Type: continuous variable.

(9) Maternal weight gain (kg) at early gestation

- This outcome will be calculated as body weight at around 20 gestational weeks minus weight before pregnancy (self-reported).
- Time Frame: from pre pregnancy to 20 gestational weeks.
- Type: continuous variable.

(10) Occurrence of gestational hypertension

- Pregnant participants will undergo blood pressure measurements at each prenatal visit. Gestational hypertension will be diagnosed if systolic blood pressure ≥140 and/or diastolic blood pressure ≥90 mmHg, based on the recommendations of the International Society for the Study of Hypertension in Pregnancy (ISSHP).
- Time Frame: during pregnancy.
- Type: binary variable (Yes / No).

(11) Maternal body weight (kg) at 6 months after delivery (Continuous variable)

(12) Maternal HbA1c (%) at 6 months after delivery (Continuous variable)

(13) Infant body weight (g) at birth, 42 day, 3 months, and 6 months (Continuous, repeated measurements)

(14) Infant length (cm) at birth, 42 day, 3 months, and 6 months (Continuous, repeated measurements)

(15) Infant body mass index (kg/m^2) at birth, 42 day, 3 months, and 6 months (Continuous, repeated measurements)

(16) Medical cost that relates with fetus and birth defects during pregnancy and after birth (Time Frame: from confirmation of pregnancy to one year old after birth) (Continuous variable)

2.2.3. Case ascertainment and case definitions

(1) Fetus and Birth defects

The examinations for detecting fetus and birth defects will follow the national routine usual care protocols. All fetus and birth defects identified from the 12 gestational weeks to 28 days after birth will be included. In consistent with the national birth defects surveillance policy, a total of 24 types of defects are defined (details of types of defects types see Appendix 1). Fetus defects will be detected by routine screening programs including prenatal DS screening, NT examination during gestational week 11 to 15, and ultrasound image examination during the second trimester (18-20 gestational week); and the number and type of birth defects after childbirth will be diagnosed and recorded according to national routine professional clinical examinations.

DS, caused by the trisomy, translocation, or partial trisomy of chromosome 21, is the most common genetic cause of intellectual disability. DS will be diagnosed by neonatologist. Neural tube defects including spina bifida and hypospadias, will be diagnosed by ultrasound examination and pediatric neurosurgeon.

CHD will be diagnosed by echocardiographic confirmation based on the local CHD screening policy (pulse oximetry plus cardiac murmurs). Hydrocephalus, digestive tract malformations, urinary malformations and other defects also will also be diagnosed either by ultrasound or some other specific diagnosis methods. From June 2016, a neonatal congenital heart defects screening policy has been issued in Shanghai city, the screening and diagnosis of CHD in neonates is embedded in this program following the routine protocols.

(2) Abortion

Abortion is defined as pregnancy suspension including miscarriage and artificial abortion. Detailed information for abortions will be collected, including time of abortion, reasons for abortion especially and antenatal information before abortion.

(3) Pregnancy complications

Pregnancy complications are a composite of gestational severe adverse events including maternal problems and fetal and placental problems. Maternal problems defined in this study include gestational diabetes, hyperemesis gravidarum, pelvic girdle pain, high blood pressure, deep vein thrombosis, anemia, infection, peripartum cardiomyopathy and hypothyroidism; Fetal and placental problems are defined as ectopic pregnancy, miscarriage, placental abruption, placenta praevia, placenta accrete, multiple pregnancies, vertically transmitted infection, intrauterine bleeding.

3. STUDY DESIGN

3.1. Design

This is a definitive, single-blind cluster randomized controlled - superiority trial embedded within the primary care system to assess the real-world effectiveness of a package of tertiary interventions before and during gestation in reducing the incidence of fetus and birth defects compared with the standard usual perinatal care. We will

recruit prepare-for-pregnancy women and their husbands through the pre-pregnancy eugenics examination clinic. Blood samples will be collected for examing RBC folate concentrations at enrollment. According to the randomization plan, participants recruited from the same community health care centre (cluster) will receive the same intervention, and clusters allocated to the two arms will receive the standard care or a serious experimental care procedure. The methods and protocols for measuring the primary outcome and the secondary outcomes in the two arms will follow the same clinical routines.

3.2. Trial Sites

The trial will recruit 22 community-based health care centers (clusters) where pre-pregnancy eugenics examination is facilitated in Shanghai: Jiuting, Fangsong, Yongfeng, Sheshan, Yueyang, Zhongshan, Xinbang, Xinqiao, Sijing, Yexie, Dongjing, Xiaokunshan, Qibao, Meilong, Pujiang, Hongqiao, Zhuanqiao, Jiangchuan, Wujing, Xinzhuang, Gumei and Maqiao community.

3.3. Interventions

Experimental interventions: usual care plus additional procedures

- **Before conception:**

  Participants will be examined for serum and RBC folate concentrations at enrollment to identify those with insufficient RBC folate concentration (<400ng/mL). Based on reviewing of folate supplement status of from the participant-administered questionnaire, a trained nurse will call the participant to inform them the RBC folate concentration and folate metabolism ability evaluated by one's genotypes of MTHFR C677T variant, and provide individualized folic acid supplementary instructions to improve folate status before get pregnant. A second RBC folate test is booked.

- **During gestation -- Green channels for confirmation of fetus defects and a following genetic and clinical consulting:**

In the both arms, fetus and infants with screen positive of fetus defect(s) will be referred via a green channel to approach an authorized tertiary obstetric hospital to achieve timely further routine comprehensive assessments and confirmative diagnosis. Then the participants with established diagnosis of fetus defects will be referred to a multi-discipline paediatric expert team for a deep consultation about possible postnatal treatments in the future to avoid unnecessary abortions.

- **After delivery:**

  The participants from the experimental arm will additionally provide timely referral to a prespecified team of pediatrician experts for timely medical treatment or scheduled follow-up.

Standard tertiary interventions of birth defects (control arm, usual care):

Couples eligible for reproductive policy are entitled to routine health cares including general health cares (health education, medical history inquiry, physical examinations, consulting guidance and pregnancy outcome follow-up) and medical examinations (laboratory examinations, virus screenings and image examinations). But nutrients status is not included in these examinations, such as folate, vitamin B12 and microelements, etc. Regular antenatal cares are required, such as deformity screening by ultrasound. Routine neonatal screenings are conducted to diagnose infant with birth defect(s) timely.

Intervention procedures of FOID study

	Experimental arm Procedures in addition to the usual care	**Control arm** Usual care
Before conception		
Preconception at enrollment	RBC folate test Inform and instruction about folic acid supplementation via phone interview	Routine physical examination and routine blood tests
During gestation		

1st antenatal visit	RBC folate test	Routine clinical examinations
15 gestational weeks		Neuro-tube defects screening
18-20 gestational week	Provided clinical and genetic consulting regarding the identified defects provided by a prespecified team of fetus and pediatrician experts	Ultrasound screen for fetus defects, referred for further confirmation diagnosis
24 gestational weeks	NO	OGTT test for screening gestational diabetes
Late gestation about 35 gestational weeks	NO	Blood test Blood routines Kidney and liver function
After delivery		
Birth	+ Neonatal CHD screening followed by ECHO for congenital heart defects diagnosis + Refer to surgery or scheduled follow-up	Recorded observed defects

3.4. Randomization and Blind

Randomization will be performed based on recruited clusters. Stata 15 was used for randomisation code generation. Twenty-two community health care centres (clusters) will be allocated to an experimental arm or a control arm with 1:1 ratio using permuted block randomisation method with block size of 4. Randomisation list will be generated by an independent statistician.

The study is single blinded, the participants from the either arm will not know the medical care provided by the opposite arm. The study investigators who are involved in recruitment, the laboratory staff, and the dose-adjusting investigator (who was not involved in participant care) will remain blinded, since all the study outcomes will occur and observed at maternity hospitals. The investigator involved in intervention

and the statistician was not blinded, as unbinding was required to carry out statistical procedures.

The fetus and birth defects observers were maternity hospital ultrasound department and neonatologists and do their work according to the usual protocol and the data of outcomes were retrieved from the medical records. Therefore, all outcome assessments will be masked. The trial statistician will also be blinded regarding the treatment code when he develops the statistical analysis plan and writes the statistical programmers, which will be validated and completed using dummy randomisation codes. The actual allocation will only be provided to the study team after lock of the database.

3.5. Sample Size

Sample size calculation was performed based on the primary outcome- the incidence of a composite rate of fetus, still birth, or birth defects. The proportion of primary endpoint was conservatively estimated to be 8% in the control arm according to two recent observational studies[6, 7] and 4% in the experimental arm (or 50% reduction) based on the findings of a meta-analysis of randomized intervention trials of folic acid supplementation intervention in folate-deficient population[8] . With cluster size of 100 subjects per centre, K coefficient of 0.23, a power of 85 % and a 5 % level of type I error, 22 clusters (2200 patients or 1100 in each arm) will be required. To allow for a 10 % dropout rate, at least 111 subjects (2446 patients or 1223 in each arm) will be recruited in each centre.

4. ANALYSIS POPULATIONS

4.1. Study Population Data Sets

Two study populations will be considered in the analysis as follows:

- **Intent-to-Treat (ITT) population**

ITT population will be defined as the pregnant women among preparing-for-pregnancy couples that were recruited since November 1, 2018, from the randomized clusters and finally get pregnant. Participants will be excluded from the ITT analysis if the primary outcome is missing, forming a modified ITT population (mITT).

- **Per protocol (PP) population**

  PP population will only include participants who received the treatments according to the protocol. Specifically,

- Participants of the experimental group: will be considered non-compliant with the preconception intervention if their RBC folate levels do not reach the target level (400 ng/ml) at the 1st antenatal visit.
- Participants in the control group received the usual care only, no non-complaints will be defined.
- Participants in both groups who do not have RBC folate levels measured at the 1st antenatal visit will not be included in the PP population.

4.2. Analysis Close Date

The analysis close date will be the date on which the last participant completed follow-up to achieve the primary outcome and until age of six months.

4.3. Data Cleaning

The data will then be checked to ensure that there are no erroneous entries and that all missing data is properly coded. Any changes will be made on the ACCESS database. After the database is cleaned and locked, it will be provided to the statistics team for analyses.

5. STATISTICAL ANALYSES

5.1. Primary Outcome Analysis

5.1.1. ITT analysis of the primary outcome - the primary analysis

The primary outcome is a binary composite outcome: occurrence of stillbirth, fetal, or birth defects identified from the 12 gestational weeks until 28 days after birth and will be summarised using number (%) of subjects with an event by treatment group.

The primary outcome will be analysed using a generalised estimating equation (GEE) model with treatment as the independent variable and center as the cluster effect. Binomial distribution, log link function and exchangeable working correlation matrix will be used in the GEE model, from which the treatment effect will be estimated as the risk ratio between the experimental arm and the control arm together with 95% confidence interval (CI). Main conclusion will be drawn from the GEE model based on ITT population.

In case the above GEE model does not converge, the following models will be fitted sequentially until convergence is achieved:

- GEE model with Poisson distribution;
- GEE model with Negative Binomial distribution;
- GEE model with Binomial distribution, logit link function and exchangeable working correlation matrix, from which odds ratio will be converted into risk ratio using the mathematical relationship between OR and RR as suggested by Localio, Margolis, and Berlin (2007).

In addition, GEE model with Binomial distribution, identity link function and exchangeable working correlation matrix will be used to calculate the risk difference between the experimental arm and the control arm together with 95% CIs. If the model does not converge, GEE model with Poisson distribution and Negative Binomial distribution will be subsequently fitted.

The above analysis will be implemented using SAS PROC GENMOD.

5.1.2. Sensitivity analyses of the primary outcome

- **PP analysis of the primary outcome**

  The same analysis of the primary outcome will be performed in the PP population using a GEE model (described in 5.1.1).

- **Covariate adjusted analysis of the primary outcome**

A GEE model used for the analysis of the primary outcome in Section 5.1.1 will also be used for covariate adjusted analysis using inversed probability of treatment weighting (IPTW) approach. The covariates will include age (continuous variable), ethnic (binary variable), body mass index (continuous variable), adverse pregnant history (binary variable), parity (continuous variable), exposure to smoking (mother or father's smoking status, binary variable), total cholesterol (continuous variable), and RBC folate level (continuous variable) at baseline to estimate adjusted risk ratio and 95% CI. Covariate adjusted analysis will be performed on both ITT and PP populations.

We will first calculate a propensity score with treatment as the dependent variable (1 for the experimental group and 0 for the reference group), and all covariates listed above as independent variables through a logistic regression model with a random effect for the center, and then perform an IPTW analysis (weighted GEE model). The adjusted risk ratio between the two groups and its two-sided 95%CI will be estimated.

5.1.3. Subgroup analysis of the primary outcome

A GEE model used for the analysis of the primary outcome in Section 5.1.1 will also be used for subgroup analysis of the primary outcome.

Subgroup analyses will be performed on age ($<$ or $\geq$ 35), ethnic (Han ethnicity or other

ethnicities), BMI (< or ≥ 24kg/m^2), presence of adverse pregnant history (Yes, No), exposure to smoking (Yes, No), total cholesterol (< or ≥ 5.2 mmol/L), RBC folate status at baseline (< or ≥ 400 ng/mL). Subgroup analyses may also base on categorized variables of the above continuous variable based on median value to make sample sizes of the categories comparable to achieve greater power.

Assessment of the homogeneity of treatment effect by a subgroup variable will be conducted by a GEE with the treatment, subgroup variable, and their interaction term as predictors, center as cluster, and the *P*-value presented for the interaction term.

Subgroup analysis will be performed on both ITT and PP populations.

5.2. Secondary Outcome Analysis

All secondary outcomes will be analysed as a superiority design and two-sided 95% CIs for the treatment differences in these outcomes between the two arms will be calculated and presented. Secondary outcome analyses will be based on the ITT population unless specified.

5.2.1. Analysis of binary secondary outcomes

Binary secondary outcomes will be analyzed with the same strategy and method that is used for the analysis of the primary outcome. GEE model will be performed with treatment as the only predictor and center as cluster effect using binomial distribution and log link function, from which the risk ratio between intervention and the control arm together with 95% CI will be derived.

Non-convergence will be dealt with using the same strategy as for the primary outcome analysis in Section 5.1.1

5.2.2. Analysis of continuous secondary outcomes

The continuous outcome will be summarised using number of subjects (n), mean, standard deviation (SD), minimum, and maximum by intervention group, and will be analysed by a GEE model with treatment as fixed effect, centre as cluster effect. The distribution will be normal distribution and link function is identity. Mean differences of outcome with their two-sided 95% CI between the two arms will be derived from the GEE model.

For repeated measurement variables, such as infant body weight at different visit, or maternal body weight during gestation, will be analyzed using linear mixed model with treatment, visit, interaction between treatment and visit as fixed effects, baseline measurement as covariate, cluster, and subject as random effects. SAS PROC GLMMIX will be used the data analysis.

5.2.3. Analysis of count secondary outcomes

The count outcome will be summarised using number of subjects (n), number of events, incidence rate (events per 100 person-years) by intervention group, and will be analysed by a GEE model with treatment as fixed effect, centre as cluster effect, and follow-up time as offset. The distribution will be Poisson distribution and link function is log. Incidence rate ratio with their two-sided 95% CI between the two arms will be derived from the GEE model.

5.3. Cost-effectiveness Analysis

We will calculate the total cost of each subject including cost of treatment, nursing, and rehabilitation from recruitment to the follow-up endpoint. The differences of total cost will be compared between the two arms using the same methods described above by treating which as a continuous variable. The total cost of intervention arm included folic

acid supplements and RBC folate examinations before conception. All the cost is settled in RMB. Further comprehensive cost-effectiveness analysis will be described by a separated SAP.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES

STATA® (version 15.0) will be used to perform all data analyses and generate the majority of data displays. SAS or SPSS or S-Plus or R may also be used for some data analyses and generating statistical graphs.

6.1. Reporting Guidelines

We will follow the Consolidated Standards of Reporting Trials (CONSORT) 2010 statement: updated guidelines for reporting cluster randomised trials (http://www.consort-statement.org/).

6.2. Participant Disposition and Flow Chart

A flow chart will be drawn up showing the number of participants screened, enrolled, and followed-up in each study arm, and the number contributing to the ITT and PP analyses.

The number screened and not enrolled and the reasons for non-enrolment will be reported, as well as the number and reasons of participants who were lost for follow up, or who were withdrawn from study for safety reason, or who crossed-over between study arms, or because of other reasons, et al.

A list of major protocol deviations will be presented after being unblind.

6.3. Data Summaries

Continuous variables will be summarized according to number of subjects with non-missing data (n), mean, standard deviation (SD), median, minimum, and maximum. The confidence intervals will be reported on summaries of continuous effectiveness variables.

Categorical variables will be summarized according to the absolute frequency and percentage of subjects (%) in each category level. The denominator for the percentages is the number of subjects in the treatment arm with data available, unless noted otherwise.

6.4. Graphical Displays

Mean values for some continuous outcomes by treatment and visit will be plotted.

6.5. Multiplicity

Multiplicity adjustment will not apply to the primary and secondary outcome analyses.

6.6. Missing Data

Missing baseline covariates will be imputed using simple imputation methods in the covariate adjusted analysis based on the covariate distributions, should the missing values for a particular covariate be less than 5%. For a continuous variable, missing values will be imputed from random values from a normal distribution with mean and SD calculated from the available sample. For a categorical variable, missing values will be imputed from random values from a uniform distribution with probabilities P_1, P_2, …, and P_k from the sample. Seed for the imputation will be set as 128.

If the missing values for a covariate are ≥5% then they will be imputed using Markov chain Monte Carlo (MCMC) methods using SAS PROC MI.

7. TABLE CONTENTS

Table 1: Summary statistics of baseline information.

Variable	Statistics	Treatment A (N=)	Treatment B (N=)	All (N=)
Age (year)	n			
	Mean			
	SD			
	<35			
	⩾35			
Ethnicity	n			
	Han			
	Other			
Preconception Body Mass Index, kg/m^2	n			
	Mean			
	SD			
	<24 kg/m^2			
	⩾2 4kg/m^2			
Parity, time	n			
	1			
	⩾2			
Presence of adverse pregnant history	n			
	Yes			
	No			
Exposure to smoking	n			
	Yes			
	No			

Table 2： Summary statistics of compliance of participants.

Variable	Statistics	Treatment A (N=)	Treatment B (N=)	All (N=)
RBC folate concentration before pregnancy	n			
	Mean			
	SD			
	Median			
	IOR			
<400 ng/ml	n(%)			
⩾400 ng/ml	n(%)			
RBC folate concentration at the first antenatal visit	n			
	Mean			
	SD			
	Median			
	IQR			
<400 ng/ml	n(%)			
⩾400 ng/ml	n(%)			

Table 3: Summary statistics of primary and secondary outcomes estimated by GEE model.

	N(%)		GEE model analysis	
Primary and secondary binary outcomes	Treatment A (N=)	Treatment B (N=)	Risk Ratio (95%CI)	p-value
Occurrence of fetus defects, still birth, and birth defects				
Occurrence of fetal defects				
Occurrence of birth defects				
Occurrence of abortion that related with congenital defects				
Occurrence of gestational hypertension				
Occurrence of maternal gestational diabetes				

Table 4: Summary statistics of secondary repeated measured outcomes estimated by linear mixed model.

		n, mean (SD)		Mixed model analysis		
Secondary repeated measured outcomes	Visit	Treatment A (N=)	Treatment B (N=)	A vs. B Mean difference（95%CI）	p-value	P-value from Treatment*time interaction test
Baby body weight	Birth					
	42 days					
	3 months					
	6 months					
Maternal body weight during gestation	Periconception Gestational week 20					
	At delivery					

8. REFERENCE

1. Berry RJ, Li Z, Erickson JD, et al. Prevention of neural-tube defects with folic acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. N Engl J Med. 1999 Nov 11;341(20):1485-90.
2. Czeizel AE, Dudás I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. N Engl J Med. 1992 Dec 24;327(26):1832-5.
3. Surén P, Roth C, Bresnahan M, et al. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. JAMA. 2013 Feb 13;309(6):570-7.
4. Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Folic Acid Supplementation for the Prevention of Neural Tube Defects: US Preventive Services Task Force Recommendation Statement. JAMA. 2017 Jan 10;317(2):183-189.
5. Li M, Zhang Y, Chen X, et al. Effectiveness of community-based folate-oriented tertiary interventions on incidence of fetus and birth defects: a protocol for a single-blind cluster randomized controlled trial. BMC Pregnancy Childbirth. 2020 Aug 20;20(1):475.
6. Xuelei Yin XD, Mingzhu Yao. Analysis on the composition of 481 neonates with congenital malformation and their associated factors. j Clin Pediatr. 2008;26(3):204-8.
7. Lixin Zheng ML, Qiling Wang, Shanshan Xu, et al. Investigation on birth defects of fetus after twelve weeks pregnancy and chiIdren with 0-5 yares old in part area of Guangdong prevince chin J Fam Plann. 2017;25(2):93-7.
8. Blencowe H, Cousens S, Modell B, Lawn J. Folic acid to reduce neonatal mortality from neural tube disorders. Int J Epidemiol. 2010;39 Suppl 1:i110-21.

9. APPENDIXES

9.1. Appendix 1. Types of fetus defects and birth defects

1. Anencephalus
2. Spina bifida
3. Encephalocele
4. Congenital Hydrocephalus
5. Cleft Palate
6. Cleft Lip
7. Cleft Lip with Cleft Palate
8. Microtia (including Anotia)
9. Deformity of external ear(s) (except Microtia and Anotia)
10. Esophageal atresia or stenosis
11. Anorectal atresia (including Congenital Anorectal Malformations)
12. Hypospadia
13. Ectopocystis
14. Pes Equinovarus
15. Polydactylism
16. Syndactylia
17. Limb shortening
18. Congenital Diaphragmatic Hernia
19. Pcromphalus
20. Celoschisis
21. Conjoined Twins
22. Trisomy 21 syndrome
23. Congenital heart disease
24. Others.

9.2. Appendix 2. FOID Study variable list

Appendix 2.1 Intervention trial variable list

NO	**variable**	
	Pre-pregnancy	
1	hospital card number	T
2	Recruited number	T
3	Recruited date	D
4	Female name	T
5	Female height	N
6	Female weight	N
7	Female age	N
8	Male name	T
9	Male height	N
10	Male weight	N
11	Male age	N
12	Tel No	T
13	Community (basis for grouping)	N
14	Nutrient Interventions	B (1=intervention 0=control)
	Nutrient first test	
15	Nutrient first test date	D
16	Serum folate	N
17	Red blood cell folate	N
18	Serum ferritin	N
19	VD(Vitamin D)	N
20	HCY(homocysteine)	N
21	Vitamin B12	N
22	LDL low density lipoprotein cholesterol	N
23	HDL high density lipoprotein cholesterol	N
24	TG total cholesterol	N
25	TC triglyceride	N
26	FBG(fasting blood-glucose)	N
27	Genotyping (BHMT (rs3733890), CBS (rs2850144, rs2851391), FIGN (rs2119289), MTHFD1 (rs2236225), MTHFR (rs1801131, rs1801133, rs3737965), MRT (rs1805087, rs28372871, rs1131450), MTRR (rs1801394, rs326119), RFC1 (rs1051266), and SHMT (rs1979277))	B
28	Clinical information (see Table 3)	-
29	Insufficient serum folate	B
30	Insufficient RBC folate	B

	Nutrient repetition measurement	
29	Nutrient repetition measurement date	D
30	Serum folate	N
31	Red blood cell folate	N
32	HCY (Homocysteine)	N
	Pregnancy	
33	Ultrasound image screen in mid-gestation	T
34	Report detail (positive)	B
35	Confirm image diagnosis	T
36	Therapeutic plan	T
37	birth defect diagnosis	T
38	Clinical information (see Table 2)	-

Note: T, text variable; D, The date type; N, Continuous variable; B, binary variable.

Appendix 2.2 Pregnant variable list in routine system

NO	variable	
	Basic information	
1	hospital card number	T
2	inpatient number	T
3	name	T
4	age	N
5	Pregnant times	N
6	Delivery times	N
7	last menstrual period	D
8	Gestational week at the first visit	N
9	The first visit date	D
10	Height	N
11	Weight	N
12	Systolic blood pressure at the first visit	N
13	Diastolic blood pressure at the first visit	N
14	Occupation	T
15	Education	T
	Antenatal care record	
16	Weight at each antenatal care	N
17	Systolic blood pressure at each antenatal care	N
18	Diastolic blood pressure at each antenatal care	N
19	Gestational week at each antenatal care	N
20	Antenatal care date	D
	Lab data	
21	Cytomegalovirus	N
22	Cytomegalovirus date	D
23	Rubella virus	N
24	Rubella virus date	D
25	Toxoplasmosis	N
26	Toxoplasmosis date	D
27	Syphilis screening	N
28	Syphilis screening date	D
29	Fasting blood-glucose	N
30	Fasting blood-glucose date	D
31	HCT(hematokrit)	N
32	HCT(hematokrit) date	D
33	Serum folate	N
34	Serum folate date	D
35	HCY(homocysteine)	N
36	HCY(homocysteine) date	D

37	OGTT 0 hours	N
38	OGTT 1 hours	N
39	OGTT 2 hours	N
40	OGTT date	D
41	Triglyceride	N
42	Triglyceride date	D
43	Total cholesterol	N
44	Total cholesterol date	D
45	Hemoglobin date	N
46	Hemoglobin date	D
	Delivery date	
47	Gestational week at delivery	N
48	Delivery mode	T
49	Birth weight	N
50	Birth weight(second baby)	N
51	Systolic blood pressure at delivery	N
52	Diastolic blood pressure at delivery	N
53	Apgar scoring	N
54	Delivery date	D
55	Birth defect records	T
56	Weight blood pressure at delivery	N

Note: The data will be extracted from maternal clinic antenatal medical record system.

T, text variable; D, The date type; N, Continuous variable; B, binary variable.

Appendix 2.3 Pre-pregnant variable list in routine system

	variable	
	Basic information	
1	Wife id	T
2	Husband nation	T
3	Husband age	N
4	Husband education	T
5	Husband id	T
6	Husband occupation	T
7	Wife nation	T
8	Wife age	N
9	Wife education	T
10	Wife occupation	T
11	Tel no	T
12	Mobile phone No	T
	Medical history	
13	Female anemia	B
14	Female EH	B
15	Female heart disease	B
16	Female DM	B
17	Female epilepsia	B
18	Female thyroid disease	B
19	Female CGN	B
20	Female mental disease	B
21	Female tumour	B
22	Female TB	B
23	Female HBV	B
24	Female VD	B
25	Male anemia	B
26	Male EH	B
27	Male heart disease	B
28	Male DM	B
29	Male epilepsia	B
30	Male thyroid disease	B
31	Male CGN	B
32	Male mental disease	B
33	Male tumour	B
34	Male TB	B
35	Male HBV	B
36	Male VD	B
	Vaccine	
37	Female rubella vaccine	B
38	Female hepB vaccine	B

39	Male hepB vaccine	B
	Durg	
40	Female current medicine	B
41	Female medicine name	B
42	Male current medicine	B
43	Male medicine name	B
	Childbearing history	
44	Birth history	B
45	Pregnancy times	B
46	Live birth	B
47	Dead fetus	B
48	Dead birth	B
49	Term delivery	B
50	Premature delivery	B
51	Natural abortion	B
52	Abactio	B
53	Children number	B
54	Birth defect	B
55	Defect type	B
56	Menarche age	B
57	Period menstruation	B
58	Menstrual cycle	B
59	Menstrual quantity	B
60	LMP	D
	Family history of disease	
61	Female family history thalassemia	B
62	Female family history albinism	B
63	Female family history favism	B
64	Female family history hemophilia	B
65	Female family history CHD	B
66	Female family history DS	B
67	Female family history openNTDs	B
68	Female family history DM	B
69	Female family history dysnoesia	B
70	Female family history daysaudia	B
71	Female family history viaual disorder	B
72	Female family history neuropsychiatric	B
73	Female family history other birthdefects	B
74	Female family history fetal death	B
75	Female family history intermarry	B
76	Female family history relations	B
77	Male family history thalassemia	B
78	Male family history albinism	B

79	Male family history favism	B
80	Male family history hemophilia	B
81	Male family history CHD	B
82	Male family history DS	B
83	Male family history openNTDs	B
84	Male family history DM	B
85	Male family history dysnoesia	B
86	Male family history dysaudia	B
87	Male family history viaual disorder	B
88	Male family history neuropsychiatric	B
89	Male family history other birth defects	B
90	Male family history fetal death	B
91	Male family history intermarry	B
92	Male family history relations	B
	Anthroposomatology	
93	Female height	N
94	Female weight	N
95	Female BMI	N
96	Female heart rate	N
97	Female SBP	N
98	Female SDP	N
99	Male height	N
100	Male weight	N
101	Male BMI	N
102	Male heart rate	N
103	Male SBP	N
104	Male SDP	N
	Lab data	
105	Leucorrhea check	N
106	Clue cell	N
107	Monilia infection	N
108	Trichomomas	N
109	Cleanness	N
110	Whiff test	N
111	PH	N
112	Wom blood analysis	N
113	Female hb	N
114	Female wbc	N
115	Female rbc	N
116	Wom urine test	N
117	Female ABO	N
118	Female Rh	N
119	Female GLU	N

120	Female GLU levels	N
121	Female NG	N
122	Female chlamydia	N
123	Female syphilis	N
124	Female HIV	N
125	Female ALT	N
126	Female ALT levels	N
127	Female HBs-Ag	N
128	Female HBs-Ab	N
129	Female HBe-Ag	N
130	Female HBe-Ab	N
131	Female HBc-Ab	N
132	Female HCV-Ab	N
133	Female CMV IgM	N
134	Female CMV IgG	N
135	Female RV IgM	N
136	Female RV IgG	N
137	Female TOX IgM	N
138	Female TOX IgG	N
139	Male blood analysis	N
140	Male hb	N
141	Male wbc	N
142	Male rbc	N

Note: The data will be extracted the preconception care electronic data system T, text variable; D, The date type; N, Continuous variable; B, binary variable.