European Journal of Epidemiology
https://doi.org/10.1007/s10654-021-00831-8

COHORT PROFILE

The China birth cohort study (CBCS)

Wentao Yue[1] · Enjie Zhang[1] · Ruixia Liu[1] · Yue Zhang[1] · Chengrong Wang[1] · Shen Gao[1] · Shaofei Su[1] · Xiao Gao[1] · Qingqing Wu[2] · Xiaokui Yang[3] · Aris T. Papageorghiou[2,4] · Chenghong Yin[1]

Received: 13 July 2020 / Accepted: 13 December 2021

Abstract

The China birth cohort study (CBCS) is a prospective longitudinal, mega-cohort study and the first national-based birth cohort study, aiming to establish a birth cohort covering representative geographical areas of the whole of China to investigate risk factors for birth defects and develop strategies for their reduction. Pregnant women who are of Chinese nationality, are 6–13^{+6} weeks of gestation, plan to attend the routine antenatal examination and deliver in the study site, and give their informed, written consent are eligible to participate in this study. All participants are followed-up through an in-person interview at 20–23^{+6} weeks and again at 28–33^{+6} weeks of gestation, and at delivery, respectively. CBCS has been divided into three phases from 20th November 2017 to 31st December 2021, and the first two phases have now been completed on 29th February 2020, enrolling 120 377 eligible pregnant women during this period. During the same period a total of 40 837 participants had been followed up to the end of pregnancy. Study recruitment will continue until December 2021 to achieve the target of 500 000 participants. Meanwhile, biological samples including peripheral blood, amniocytes, cord blood, placenta, or umbilical cord tissue have been collected from participants according to various conditions. The incidence of birth defects in this group is 2.5% and congenital heart disease is the most common type of birth defect seen so far. A website is in the advanced stages of planning, to allow seamless data transfer and facilitate collaboration with groups around the world.

Keywords Cohort study · Birth defects · Pregnancy · China

✉ Aris T. Papageorghiou
aris.papageorghiou@wrh.ox.ac.uk

✉ Chenghong Yin
modscn@126.com; yincch@ccmu.edu.cn

Wentao Yue
yuewt@ccmu.edu.cn

Enjie Zhang
zhangenjie@mail.ccmu.edu.cn

Ruixia Liu
liuruixia@ccmu.edu.cn

Yue Zhang
zhangyue0517@ccmu.edu.cn

Chengrong Wang
chrwang@ccmu.edu.cn

Shen Gao
shengao@mail.ccmu.edu.cn

Shaofei Su
movingsu@mail.ccmu.edu.cn

Xiao Gao
gaoxiao1206@163.com

Qingqing Wu
qingqingwu@ccmu.edu.cn

Xiaokui Yang
yangxiaokui@ccmu.edu.cn

[1] Department of Central Laboratory, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing, People's Republic of China

[2] Department of Ultrasound, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing, People's Republic of China

[3] Department of Reproductive Medicine, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing, People's Republic of China

[4] Nuffield Department of Women's and Reproductive Health, University of Oxford, John Radcliffe Hospital, Oxford, UK

Introduction

Birth defects are an important cause of infant mortality and lifelong disability, which result in tremendous burden to families and the society. According to the 2010 Global Burden of Disease (GBD) study, 6.4% of neonatal infant deaths are attributed to birth defects, which ranked 5th among all causes of death [1]. This burden appears not to be evenly distributed, rather it has been reported that the prevalence of all birth defects in live births ranges from a high of 8.2% to a low of 4.0% worldwide. In addition to lethal birth defects, it has been estimated that at least 3.2 million survivors may suffer from significant disability for life [2]. It is important to note that the impact of birth defects is particularly severe in middle- and low-income countries including China. According to the report of Chinese Birth Defect prevention in 2012, there are about 900,000 annual births that are affected by birth defects, accounting for 5.6% of all births [3]. At the same time, given the rapid economic development, industrialization and urbanization in China and beyond, concerns have been raised in relation to environmental pollution as a modifiable influence on pregnant women and, by extension, on the prevalence of birth defects [4]. Hence, a better understanding of demographic distributions and risk factors associated with birth defects is urgently needed to provide evidence for etiology and prevention. We believe there is an important need to determine these exposotypes at the population level as they may help us to understand how exposures affect the occurrence of birth defects at the individual and systems level, and can therefore lead to determining cause, nese t, eold wisde.ept'pritons. Hdefe39999.9mbst i.8(the y8re)1.6(u.5 Tc--.045 0 T25 Td[(longituda bimain hce oypi18def)34(esit)-7.our)15(ee evidence comes from high income settings in the USA and Europe, with studies rather insufficient and limited in Chinese populations [5–8].

The China birth cohort study (CBCS) is a prospective, longitudinal, mega-cohort study ultimately aiming at prevention of birth defects in China. The CBCS aims to establish a birth (d).i
s

pre-planned interim assessment of all processes for the study. This is the subject of this report, and we were particularly interested to examine follow-up rates, and establish whether recruitment of more sites during the final phase (from 1st March 2020 to 31st December 2021) is required, thus ensuring that the study can be completed on time.

The first two phases have now been completed, with phase 2 finishing, as planned, on 29th February 2020. On that date we had enrolled 120 377 eligible women in early pregnancy (less than 14 weeks of gestation at enrolment), accounting for just over 24.1% of the total target. This recruitment was completed at 38 research sites in 17 provinces, cities, autonomous regions and municipalities covering most areas of China (Fig. 1). Most of these sites are referral hospitals (3A hospitals), with a total of just over 300 000 deliveries annually.

Followed up strategies

In CBCS, women are enrolled in early pregnancy, at 6–13^{+6} weeks of gestation. At this point, all participants are asked to complete a baseline questionnaire and donate 10 ml of peripheral blood, taken before 13^{+6} weeks of gestation. Clinical laboratory measures are collected for each participant at recruitment (Supplementary table 1). The first and second follow-up visits are undertaken mid-pregnancy at 20–23^{+6} and late pregnancy at 28–33^{+6} weeks of gestation, respectively. For all participants, questionnaires are completed at these two follow-up visits by in-person interviews at their routine prenatal examination. Corresponding clinical laboratory measures are collected at both of these follow-up visits (Supplementary table 1). The third follow-up visit is undertaken after delivery. The clinical information is recorded for all the participants by trained researchers, doctors or nurses.

If the participant has a miscarriage in early pregnancy, or a pregnancy loss during mid- or late pregnancy, all clinical information will be recorded by trained researchers, doctors or nurses. If a birth defect is found at any of these stages, clinical information, including ultrasound scan information

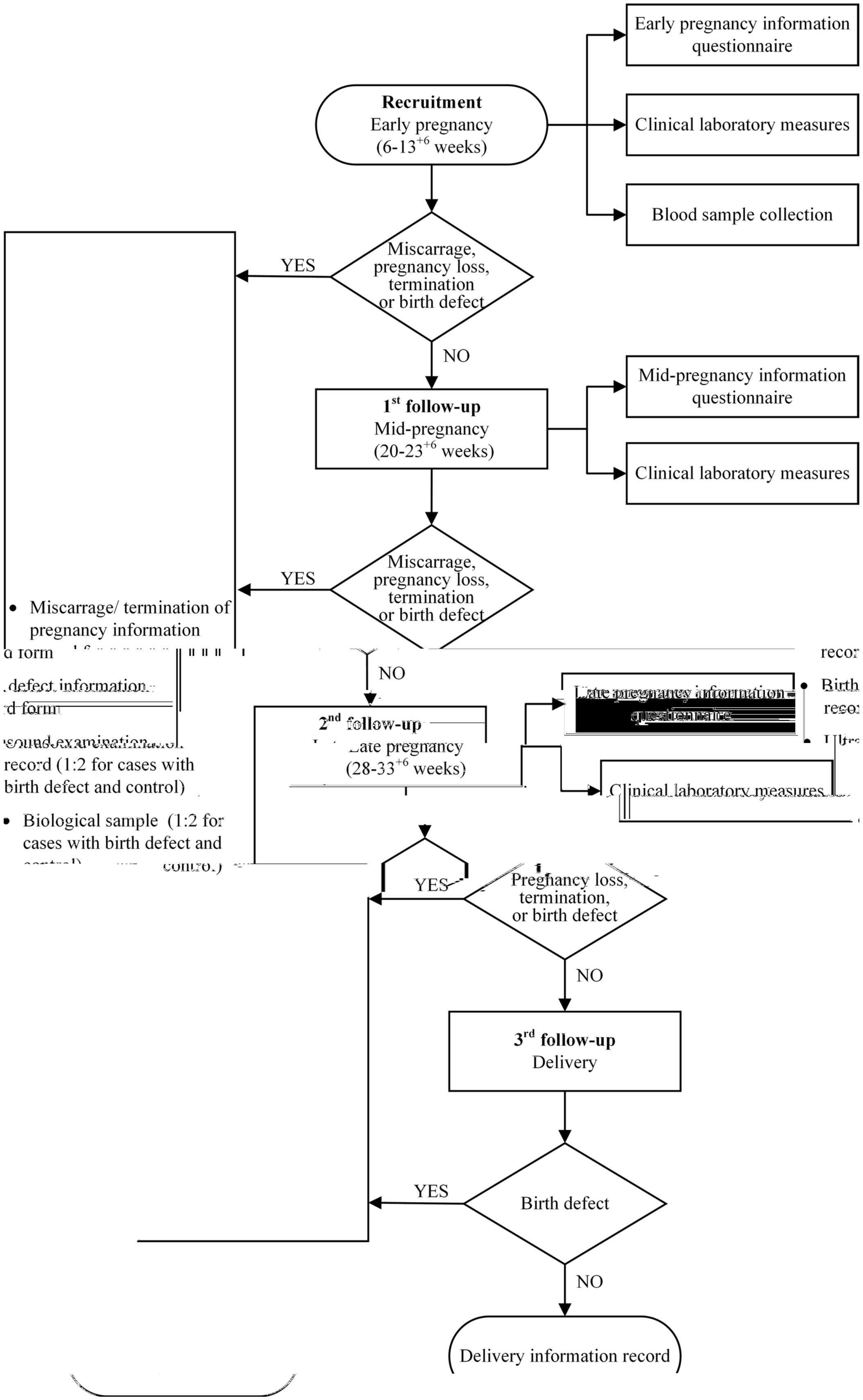

Fig. 2 Flow chart of cohort participant recruitment and follow-up

defects, occupation, exposures at home and in workplace and lifestyle habits. A food frequency questionnaire (Q2: 70 questions) is required for a representative sample of 80 000 participants. The questionnaires (Q3: 24 questions and Q4: 17 questions) used for the first two follow-up stages, cover the pregnant woman's health status as well as any prenatal diagnoses. For Q5 (14 questions) data on mode of delivery, perinatal outcomes and birth characteristics (standardised measurement of birth weight, birth length, head circumference, new-born sex, Apgar score and the information of placenta and umbilical cord) are collected at the third follow-up, according to international standards [9]. The information about fathers was embedded in Q1 completed at enrolment, which was filled in by pregnant women, including demographic characteristics, family history of birth defects, smoking status and alcohol use.

All clinical laboratory measures include routine blood tests, as well as blood biochemistry, thyroid function, coagulation function, vitamin level tests, screening for Toxoplasma, Rubella virus, Cytomegalovirus and Herpes virus, maternal serum alpha fetoprotein, oral glucose tolerance test (OGTT) and HbA1c (Supplementary Table 1). Biological samples collected at different periods during pregnancy include peripheral blood samples donated by all participants, and may include amniocytes for those women opting to have amniocentesis, tissue, placenta, cord blood or umbilical cord tissue donated by participants with pregnancy loss or birth defects, and, where appropriate, controls.

Data management

All baseline, follow-up and outcome data, as well as clinical laboratory measurements and bio-banked samples, are managed through a secure cloud-based platform, including an electronic data capturing (EDC) and a bio-bank system. This platform can combine the information from the two systems automatically and is used to establish, manage and maintain the database. Data can be collected in different ways including on mobile phone, PC or data import. Logic check functions are programmed during data input or capturing in order to avoid data entry errors and missing items, and to ensure integrity and accuracy of the data. The logic validation is not only in each questionnaire, but also operates across different questionnaires for the same woman. Accounts have different levels of permissions for research staff, database administrators and investigators. The platform issues follow-up reminders to both administrators and participants at each follow-up stage. Overall, this cloud-based platform has been designed based on the main principles of data security; ongoing quality control; the ability to share data among different research sites; ease of usability; and extensibility.

The data, including all the measurements in the questionnaire and biological information from each participating centre, are transferred into the central CBCS Database in real time when the information is submitted. Additionally, all collected data is backed up and sent to the central server once a week. The data is managed by professional data administrators and will not be exported until the CBCS steering committee agrees.

Currently, as our study is still in progress, our data cannot be shared externally. However, the data is shared within the team and among participating centres. We will consider the time of data sharing externally when recruitment and follow-up is completed. There will be an application process for use of the data through the completion of a request form which will include the proposed analysis strategy. Applications will be evaluated and once approved by the CBCS committee, data will be shared; results from such analysis must be returned to the database, to allow further sub-analysis.

Baseline characteristics

The current data are from the pre-planned interim analysis, prior to the third phase of the study, undertaken to ensure proper functioning of all processes, examine follow-up rates, establish whether recruitment of more sites is needed and highlight any problems. As the study is still ongoing, early findings are described below. Table 2 presents the current baseline characteristics of study population. A total of 120 377 pregnancy women have been enrolled, nearly half nulliparous. More than 95% of pregnancies were conceived naturally, while 4.2% used some form of assisted conception. The mean maternal age was 30.08 ± 4.27 and for paternal age it was 31.45 ± 4.93. More fathers were ≥ 36 years old (18.52%) than mothers (11.25%). The age at menarche for most women was 12–14 years, similar to the majority participants in USA, Germany and African-Americans [10–12]. Very few women smoked or consumed alcohol regularly, and

Table 1 Information collected by questionnaires, physical measurements, laboratory measurements and medical record abstraction, and biological samples collected in the China Birth Cohort study

Stages	Measurements	Methods	Biobank
Early Pregnancy (6–13^{+6} gestation weeks) Enrollment	*Demographic characteristics:* data of birth, ethnic group, education, occupation/employment, income, birthplace; *Health:* height, weight (at preconception and early pregnancy), last menstrual period, blood pressure, health status, age of menarche, menstruation status, medical history, family history of birth defects; *Reproductive status:* method of conception, number of fetuses, reproductive history, history of pregnancy complications; *Lifestyle:* smoking status, alcohol, medication and supplements (folic acid, multi-vitamins), pets exposure; *Environmental exposure:* housing characteristics, decoration, method of heating, use of air purifier, chemical exposure, use of pesticides, occupational exposure, working category, working hours; *Clinical factors and laboratory measurements;* *Diets:* eating habits of the past week, main meals in the past month including staple food, beans, vegetables, fruits, milk, meat, fish products, eggs, snacks, drinking water and soft drinks, cooking oil, dietary supplement;	Structured questionnaire	Whole blood, serum, Chorionic tissue from the fetus*
Mid-pregnancy (20–23^{+6} gestation weeks) Follow-up	Blood pressure, threatened abortion, pregnancy complications (GDM, PIH and Thyroid function); *Prenatal screening:* Down's syndrome screening results, noninvasive prenatal testing, genetic screening for deafness, amniocentesis and other interventional prenatal diagnosis; *Clinical factors and laboratory measurements;*	Structured questionnaire/Physician’s medical records	Amniocyte, cutaneous or muscular tissue from the fetus, placenta, cord blood or umbilical cord tissue*
Late Pregnancy (28–33^{+6} gestation weeks) Follow-up	Blood pressure, threatened premature labor, pregnancy complications (GDM, PIH and Thyroid function); Prenatal screening: amniocentesis and other interventional prenatal diagnosis; *Clinical factors and laboratory measurements;*	Structured questionnaire/Physician’s medical records	Amniocyte, cutaneous or muscular tissue from the fetus, placenta, cord blood or umbilical cord tissue*

Table 1 (continued)

Stages	Measurements	Methods	Biobank
Delivery Follow-up	Delivery data, delivery mode, neonatal sex, birth weight, birth length, head circumference, Apgar score, placenta size and shape, umbilical cord length;	Structured questionnaire/Physician's medical records	Cutaneous or muscular tissue from the fetus, placenta, cord blood or umbilical cord tissue*
Father information	*Demographic characteristics:* data of birth, ethnic group, education, occupation/employment, income; *Health***:** height, weight, family history of birth defects; *Lifestyle:* smoking status, alcohol;	Structured questionnaire	

*Only for cases with birth defects and controls (1:2). PIH: pregnancy induced hypertension. DM: diabetes mellitus

Peripheral blood samples are collected from all participants at enrollment, then processed and stored as whole blood and serum. Chorionic tissue from the fetus are collected from women who have a miscarriage in early pregnancy. Amniocytes are collected for those women opting to have amniocentesis during mid- and late pregnancy. Cutaneous or muscular tissue from the fetus, placenta, cord blood or umbilical cord tissue are collected from women with an induced labour or a birth defect, and cord blood or umbilical cord from controls

Main strengths and weaknesses

The CBCS has a number of strengths. Most important is the very large sample size, the detailed, comprehensive information, and multiple biological samples collected. To our knowledge, this will be the largest birth cohort in the world and will provide adequate power to investigate the causal effects of different exposures on birth defects [14]. The data collected in this study cover environment factors, genetic factors, medication exposure and medication, chronic disease, nutrition and lifestyle. Measuring this multitude of exposures in a single study, coupled with detailed follow up and outcome ascertainment mean that associations can be explored in detail. For example, the relationship between environment (atmosphere, greenness, light, noise, etc.) and teratogenesis, or the associations of the use of drugs with unclear teratogenic effect and birth defects. Moreover, extensive maternal blood samples are collected. Remarkable advances in -omics technologies, including genomics, metabolomics, proteomics and assessment of the microbiome have provided new opportunities for systematic epidemiologic research and further exploration of the mechanism of birth defects. Therefore, integrating-omics and digital technologies, and incorporating a multidisciplinary approach across the life cycle, should be most effective for understanding the factors associated with, and ultimately the prevention, of birth defects.

One weakness of this cohort is the selection bias which had been recognized. However, the main aim of this study is to explore the associations between maternal exposure and birth defects and other adverse outcomes, and the selection is balanced for maternal exposure and the outcomes since the pregnancy outcome is not known at recruitment. Another weakness of this cohort is the self-administered questionnaire at recruitment, with some data based on participant self-evaluation, including health-related variables. However, the questions are presented in simple language and have been piloted; comparison undertaken with forms completed by healthcare staff have demonstrated high concordance. Furthermore, we have collected many different types of biological samples, but we have not started to analyse these. However, genomic sequencing is planned, and we are keen to incorporate emerging new technologies.

In our study, the neonatal outcome at delivery is the final end-point, and no further follow-up of participating mothers and their children is being carried out; this is a weakness when compared to other long-standing birth cohort studies, such as the Danish National Birth Cohort (DNBC) and the Norwegian Mother and Child Cohort Study (MoBa) [14–16]. This is due to financial constraints and a study time-line that funders have implemented; it means that longer-term conditions in the infant, that are not evident at birth, may be under-ascertained. Detailed newborn assessment is being undertaken to mitigate this risk. At

Table 2 Baseline characteristics of participants in the China Birth Cohort study

Characteristics	No. of samples (Mean, 95%CI/median)	Percentage (Standard deviation/IQR)
Maternal age at pregnancy, years		
Mean age, 95%CI	30.08, 30.06–30.11	(4.27)
Median age	30.00	(27.00–33.00)
≤ 25	15354	12.75
26–30	53 717	44.62
31–35	37761	31.37
≥ 36	13545	11.25
Paternal age, years		
Mean age (SD)	31.45, 31.42–31.48	(4.93)
Median age	31.00	(28.00–34.00)
≤ 25	9888	8.21
26–30	46 744	38.83
31–35	41455	34.44
≥ 36	22290	18.52
Maternal ethnicity		
Han	112712	93.63
Others	7665	6.37
Paternal ethnicity		
Han	113988	94.69
Others	6389	5.31
Maternal education		
< College	27923	23.20
Undergraduate/College	79942	66.41
> Postgraduate or higher	12512	10.39
Paternal education		
< College	30696	25.50
Undergraduate/College	77184	64.12
> Postgraduate or higher	12497	10.38
Maternal occupation		
Unemployed	29293	24.33
Office worker	9820	8.16
Doctor or medical service personnel	4713	3.92
Banking or commerce or service staff	5970	4.96
Company staff	44145	36.67
Manufactory worker or farmer	1652	1.37
Teacher or researcher or technician	11360	9.44
Military or police	281	0.23
Others	13143	10.92
*Maternal annually income, Yuan**		
< 50 000	41195	34.22
50 000–100 000	47277	39.27
100 000–200 000	23796	19.77
200 000–400 000	6328	5.26
400 000–600 000	1192	0.99
> 600 000	589	0.49
*Family annually income, Yuan**		
< 50 000	15138	12.58
50 000–100 000	34289	28.48
100 000–200 000	39346	32.69

Table 2 (continued)

Characteristics	No. of samples (Mean, 95%CI/median)	Percentage (Standard deviation/IQR)
200 000–400 000	21067	17.50
400 000–600 000	7040	5.85
> 600 000	3497	2.91
Maternal smoking	225	0.19
Paternal smoking	40665	33.78
Maternal alcohol drinking	3950	3.28
Paternal alcohol drinking	38454	31.94
PIH		
Yes	384	0.32
No	119993	99.68
DM		
Yes	316	0.26
No	120061	99.74
Maternal biological families with birth defects	2038	1.69
Paternal biological families with birth defects	1633	1.36
Age of menarche, years		
Mean age	13.20, 13.19–13.21	(1.27)
Median age	13.00	(12.00–14.00)
< 12 years	3721	3.09
12–14 years	99492	82.65
> 14 years	17164	14.26
Pre-pregnancy BMI, kg/m^2		
Mean BMI	22.06, 22.04–22.08	(4.23)
Median BMI	21.26	(19.53–23.59)
< 18.5	16132	13.40
18.5–25	85662	71.16
25–30	14120	11.73
≥ 30	4463	3.71
Paternal BMI, kg/m^2		
Mean BMI	24.96, 24.93–24.99	(4.95)
Median BMI	24.24	(22.13–26.73)
< 18.5	3212	2.67
18.5–25	67430	56.02
25–30	39628	32.92
≥ 30	10107	8.40
Fertilization way		
Natural pregnancy	114679	95.27
Artificial insemination	637	0.53
IVF		

the same time we plan to expand the scope of disease and health assessment in future to explore the causal relationship between early life exposures and later health status. In addition, we are aiming to extend our work in order to collect additional information about the father, establish a biological sample database, and develop a long term follow-up plan covering the entire life cycle.

The increasing number of birth cohort studies over recent decades has given opportunities for projects to integrate multiple cohorts, such as the Environmental influences on Child Health Outcomes (ECHO) program which contains 84 observational cohorts and the EU Child Cohort Network which contains 19 pregnancy and childhood cohorts [17, 18]. Combining data will make it possible to identify smaller effect estimates, and search for differences in risk factors across countries. This will enable better research into causal understanding and modelling of life course health trajectories. To maximize the benefits of our research, we are already forming links to collaborate with other birth cohort studies, and our data will ultimately be made accessible to other researchers.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10654-021-00831-8.

Acknowledgements The authors are grateful to the participants in the CBCS for their support in the data collection.

Funding This work was supported by The National Key Research and Development Program of China (2016YFC1000101).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The lancet. 2012;380(9859):2095–128.
2. Christianson AL, Howson CP, Modell B: Global report on birth defects: the hidden toll of dying and disabled children: March of Dimes Birth Defects Foundation; 2006.
3. The report of Chinese Birth Defect Prevention in 2012 [http://www.gov.cn/gzdt/2012-09/12/content_2223371.htm]
4. Kan H. Globalisation and environmental health in China. The Lancet. 2014;384(9945):721–3.
5. Qiu X, Lu J-H, He J-R. Lam K-bH, Shen S-Y, Guo Y, Kuang Y-S, Yuan M-Y, Qiu L, Chen N-N: The born in Guangzhou cohort study (BIGCS). Eur J Epidemiol. 2017;32(4):337–46.
6. Tao F-B, Hao J-H, Huang K, Su P-Y, Cheng D-J, Xing X-Y, Huang Z-H, Zhang J-L, Tong S-L. Cohort profile: the China-Anhui birth cohort study. Int J Epidemiol. 2013;42(3):709–21.
7. Zhang J, Tian Y, Wang W, Ouyang F, Xu J, Yu X, Luo Z, Jiang F, Huang H, Shen X. Cohort profile: the Shanghai birth cohort. Int J Epidemiol. 2019;48(1):21–21g.
8. Zheng J-S, Liu H, Jiang J, Huang T, Wang F, Guan Y, Li D. Cohort profile: The jiaxing birth cohort in china. Int J Epidemiol. 2017;46(5):1382–1382g.
9. Villar J, Ismail LC, Victora CG, Ohuma EO, Bertino E, Altman DG, Lambert A, Papageorghiou AT, Carvalho M, Jaffer YA. International standards for newborn weight, length, and head circumference by gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. The Lancet. 2014;384(9946):857–68.
10. Dreyfus J, Lutsey P, Huxley R, Pankow J, Selvin E, Fernandez-Rhodes L, Franceschini N, Demerath E. Age at menarche and risk of type 2 diabetes among African-American and white women in the Atherosclerosis Risk in Communities (ARIC) study. Diabetologia. 2012;55(9):2371–80.
11. Saquib N, Kritz-Silverstein D, Barrett-Connor E. Age at menarche, abnormal glucose tolerance and type 2 diabetes mellitus: the Rancho Bernardo Study. Climacteric. 2005;8(1):76–82.
12. Stöckl D, Döring A, Peters A, Thorand B, Heier M, Huth C, Stöckl H, Rathmann W, Kowall B, Meisinger C. Age at menarche is associated with prediabetes and diabetes in women (aged 32–81 years) from the general population: the KORA F4 Study. Diabetologia. 2012;55(3):681–8.
13. China Statistical Yearbook 2019 [http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm]
14. Olsen J, Melbye M, Olsen SF, Sørensen TI, Aaby P, Nybo Andersen A-M, Taxbøl D, Hansen KD, Juhl M, Schow TB. The Danish National Birth Cohort-its background, structure and aim. Scandinavian journal of public health. 2001;29(4):300–7.
15. Magnus P, Birke C, Vejrup K, Haugan A, Alsaker E, Daltveit AK, Handal M, Haugen M, Høiseth G, Knudsen GP. Cohort profile update: the Norwegian mother and child cohort study (MoBa). Int J Epidemiol. 2016;45(2):382–8.
16. Magnus P, Irgens LM, Haug K, Nystad W, Skjærven R, Stoltenberg C. Cohort profile: the Norwegian mother and child cohort study (MoBa). Int J Epidemiol. 2006;35(5):1146–50.
17. Gillman MW, Blaisdell CJ. Environmental influences on child health outcomes, a research program of the NIH. Curr Opin Pediatr. 2018;30(2):260.
18. Jaddoe VW, Felix JF, Andersen A-MN, Charles M-A, Chatzi L, Corpeleijn E, Donner N, Elhakeem A, Eriksson JG, Foong R: The LifeCycle Project-EU Child Cohort Network: a federated analysis infrastructure and harmonized data of more than 250,000 children and parents. Eur J Epidemiol 2020:1–16.