



---

## Data Resource Profile

# Data Resource Profile: Xiamen registry of pregnant women and offspring (REPRESENT): a population-based, long-term follow-up database linking four major healthcare data platforms

Jing Tan,<sup>1,2</sup> Yiquan Xiong,<sup>1</sup> Yana Qi,<sup>1</sup> Chunrong Liu,<sup>1</sup> Shiyao Huang,<sup>1</sup> Guanhua Yao,<sup>3</sup> Wei Sun,<sup>3</sup> Yongyue Qian,<sup>3</sup> Lishan Ye,<sup>4</sup> Qiushi Xu,<sup>4</sup> Hui Liu,<sup>4</sup> Andy H Lee,<sup>5</sup> Lehana Thabane<sup>2</sup> and Xin Sun<sup>1\*</sup>

<sup>1</sup>Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, China, <sup>2</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada; Biostatistics Unit, St Joseph's Healthcare—Hamilton, Hamilton, Canada, <sup>3</sup>Xiamen Health Commission, Xiamen, China, <sup>4</sup>Xiamen Health and Medical Big Data Center, Xiamen, China, <sup>5</sup>School of Public Health, Curtin University, Perth, Australia

\*Corresponding author. Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu 610041, China. E-mail: sunx79@hotmail.com

Editorial decision 27 July 2020; Accepted 31 July 2020

---

## Data resource basics

### Origins of data source

As the nation with the largest population of >17 million newborns annually,<sup>1</sup> China always places maternal and child health on its priority research agenda.<sup>2</sup> During the past decades, substantial interventional works have led to a large reduction in maternal and infant mortality.<sup>3,4</sup> With the introduction of a national goal of Healthy China 2030,<sup>5</sup> the research paradigm has now shifted and expanded to the improvement of overall health, such as reduction of newborn and infant malformations, and lowering of serious pregnancy-related complications. This paradigm shift is particularly relevant to address new challenges for advanced-age mothers since the implementation of China's two-child policy.<sup>6</sup> Such challenges include the necessity to answer policy-relevant questions in a timely manner, the limited resources available to deal with rare but serious adverse birth events, and the lack of scientific evidence to causally determine the relationship

between pre-pregnancy exposures and long-term health outcomes.

The advancement of digital technology in the past decade has enabled electronic collection and storage of healthcare data in China.<sup>7,8</sup> The National Health Commission has also enforced the capturing of healthcare data for women and children, such as maternal death, under-5 mortality and birth defects.<sup>9</sup> Xiamen, a sub-provincial city of >4 million residents, is unique in the governance of healthcare data in China. Through a central coordination system, electronic health records are centrally collected from primary to tertiary care institutions within the region, whereby the same electronic health records system is adopted across all primary care institutions, and a unique pregnancy registry has been established. It is thus feasible to utilize such integrated data sources to generate trustworthy evidence on pregnancy and child care, spanning several years before and after pregnancy, and providing valuable information about mothers and their offspring. In this data source profile, we report our efforts

### Key Features

- REPRESENT is the first population-based, long-term follow-up database of pregnant women and their offspring in China that can capture information on a diversity of exposures before pregnancy as well as long-term outcomes for both mothers and infants, in addition to tracking data during gestation and delivery.
- In the past 11 years (January 2008 to March 2019), REPRESENT has accumulated data concerning 766 194 pregnancies and 765 746 newborns, with over 800 variables being documented from four healthcare data platforms in Xiamen, a sub-provincial city in southeast China that has 4.11 million residents.
- REPRESENT was primarily built upon the Maternal and Child Health Management Platform, a pregnancy registry which stores healthcare data—including antenatal care, delivery care, postpartum follow-up and childhood follow-up—from pregnant women and their offspring across all maternity departments in Xiamen.
- By using a unique identifier, REPRESENT linked the pregnancy registry with three other healthcare data platforms, which store healthcare data from 15 tertiary and 2 secondary hospitals, 39 primary healthcare facilities and 272 (of the 281) health stations in the city. The scope of the database spans from pre-pregnancy exposures to post-delivery information about health care and outcomes for mothers, and the follow-up for offspring may cover adolescence and adulthood.
- Research institutions can apply to access the registry data by submitting a formal study protocol, subject to the approval by Xiamen Health and Medical Big Data Center and Chinese Evidence-based Medicine Center ([http://www.xmyys.com.cn/news\\_info\\_204.html](http://www.xmyys.com.cn/news_info_204.html)). Ethical review and research registration are mandatory for all studies.

to develop the Xiamen registry of pregnant women and offspring (REPRESENT) that links four major healthcare data platforms in the region.

### Coverage and context

Xiamen is a coastal city located in southeast China, with a subtropical monsoon climate at an average annual temperature of 21°C. The sub-provincial city is economically administered by the central government of China, and the total population of permanent residents is 4.11 million. The city comprises 6 administrative districts, 38 sub-district units, 364 residential communities and 147 villages. Xiamen currently maintains 15 tertiary hospitals (defined as those with >500 beds), 2 secondary hospitals (between 100 and 499 beds), 39 primary healthcare facilities and 281 health stations (Figure 1).

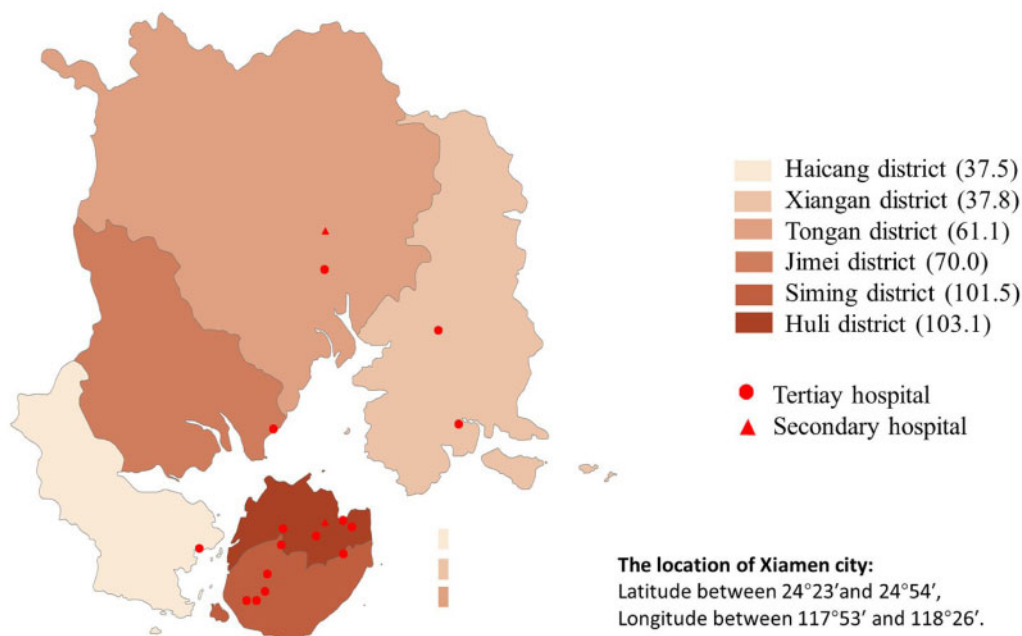
Since 2006, the government of Xiamen has developed four major healthcare data platforms (Figure 2): (i) Residents Healthcare Management Platform, which integrates data from the other three platforms and serves for making appointments and checking electronic health files by residents (Supplementary Appendix 1A, available as Supplementary data at *IJE* online); (ii) Primary Healthcare Management Platform, which collects healthcare data from 39 primary healthcare facilities and 272 (of the 281) health stations (Supplementary Appendix 1B, available as Supplementary data at *IJE* online); (iii) Electronic Healthcare Records (EHR) Platform, which collects EHR

data from 15 tertiary and 2 secondary hospitals (Supplementary Appendix 1C, available as Supplementary data at *IJE* online); and (iv) Maternal and Child Health Management Platform, which collects healthcare data from pregnant women and their offspring across all maternity departments in Xiamen.

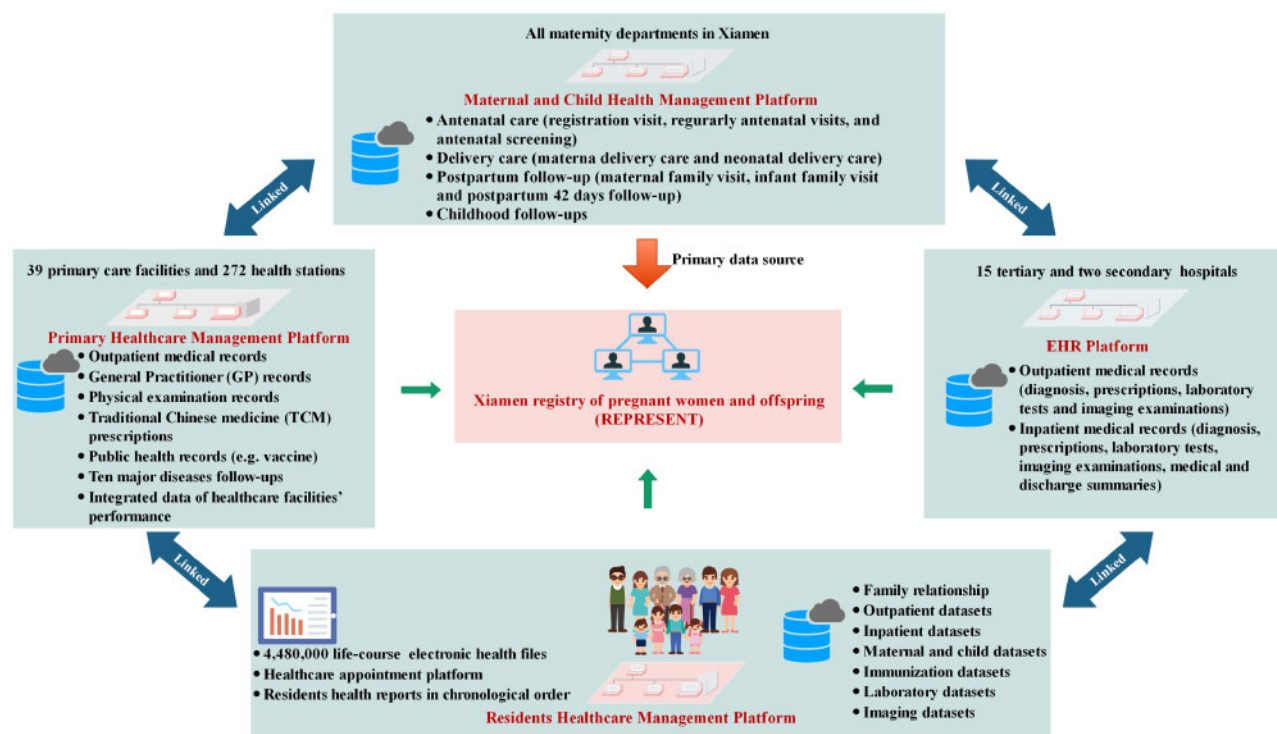
Xiamen residents are given a healthcare card with a unique Citizen Health Identification number, linked to their citizen identity number. They are required to use the healthcare card for making medical appointments, either in person at the reception desk of the healthcare institution or through the Management Platform, which has computer or mobile portals. The identification number enables the linking of data across different platforms.

Xiamen Health Commission holds primary responsibility for the development and management of the platforms, data coordination, transfer, uploading and storage, as well as quality control. The Xiamen Health and Medical Big Data Center, under the governance of Xiamen Health Commission, is responsible for technical operation of the platforms.

The establishment of REPRESENT commenced with setting up the pregnancy registry, primarily based on the Maternal and Child Health Management Platform. This registry was then linked to the three other platforms. Such data linkage enables longitudinal follow-up of pregnant women and their offspring. The scope of the database spans from pre-pregnancy exposures to post-delivery



**Figure 1** Population coverage and location of hospitals at the end of 2018 in Xiamen city. Shading represents the population coverage of permanent residents in each district (10 000), circles represent the location of tertiary hospitals, and triangles represent the location of secondary hospitals



**Figure 2** Four major healthcare data platforms in Xiamen. These databases are linked through the Citizen Health Information number (a unique ID). The REPRESENT registry comprises of data from the four databases, with the Maternal and Child Health Management Platform as the primary data source

information about health care and outcomes for mothers, and the follow-up for offspring may cover adolescence and adulthood. The current database has accumulated 11-year data ranging from 2008 to 2019, and the ongoing

accumulation of data would enable a great potential for life-course coverage.

Table 1 summarizes the sample size and duration time for each data panel and its subset categories. A total of

**Table 1** Sample size and duration time of data panel and subset in the Maternal and Child Health Management Platform

Data set	Subset	Duration time	Numbers	Values (no.)
Antenatal care	Registration visit	Jan 2008–Mar 2019	Number of pregnancies	766 194
			Number of mothers	591 592
	Regularly antenatal visits	Jan 2008–Mar 2019	Number of pregnancies	684 339
			Number of mothers	535 477
			Number of repeated visits	4 771 235
			Number of repeated visits per pregnancy (median, Q <sub>L</sub> –Q <sub>U</sub> )	7 (4–10)
	Antenatal screening	Jan 2008–Mar 2019	Number of records	3 266 906
			Number of pregnancies	371 434
			Number of mothers	352 754
Delivery care	Maternal delivery care	Jan 2008–Mar 2019	Number of pregnancies	758 237
			Number of mothers	529 424
	Neonatal delivery care	Jan 2008–Mar 2019	Number of newborns	765 746
			Number of pregnancies	755 794
			Number of mothers	526 050
			Number of records	304 946
Postpartum follow-ups	Maternal family visit	Jan 2014–Mar 2019	Number of pregnancies	229 604
			Number of mothers	204 250
			Number of records	286 994
	Infant family visit	Jan 2010–Mar 2019	Number of records	212 739
			Number of mothers	140 947
			Number of pregnancies	140 947
	Postpartum 42 days follow-up	Jan 2013–Mar 2019	Number of records	171 793
			Number of pregnancy	169 991
			Number of individuals	155 589
Childhood follow-ups	Jan 2008–Mar 2019	Number of repeated records	4 90 697	
		Number of children	828 659	

766 194 pregnancies are registered at Maternal and Child Health Management Platform, including 4 771 235 records of repeated antenatal visits and 3 266 906 records of antenatal screening. In addition, 758 237 pregnancies in maternal delivery care and 765 746 newborns in neonatal delivery care are documented.

Table 2 presents maternal and offspring characteristics at baseline and key information of the REPRESENT database. The median registrations per year was 55 765 [lower quartile (Q<sub>L</sub>)–upper quartile (Q<sub>U</sub>), 35 020–96 973], and the average maternal age was 27 (25–30) years. A total of 328 375 (46.29%) pregnancies were multipara and 245 822 (32.49%) deliveries were by caesarean section.

## Data collected

### Pregnancy registry development

In an effort to protect the health of mothers and offspring, the National Health Commission of China has implemented rigorous pregnancy management since 2011.<sup>10</sup> All the maternal departments in Xiamen have followed two guidance documents, issued by the National Health

Commission, to develop the pregnancy registry and conduct follow-ups, including Administrative Regulation on Health Care during Pregnancy and Childbirth and Good Practice for Health Care during Pregnancy and Childbirth.<sup>10</sup> Thus, pregnancy care in Xiamen follows a standard protocol (Figure 3).

After confirmation of conception by ultrasonography, a pregnant woman is registered at the Maternal and Child Health Management Platform by each clinician at the first antenatal visit, and a unique pregnancy registration number is assigned. The woman is subsequently assessed by a clinician using a maternal risk score, and her pregnancy is categorized as either low risk (without any high-risk characteristics) or high risk (with any high-risk characteristics).<sup>11</sup>

Following the initial assessment, antenatal care for low-risk pregnancies is conducted at primary or secondary healthcare facilities until 20 weeks of gestation; if any complication occurs, the woman is transferred to a tertiary hospital. Nuchal translucency examination is conducted at a tertiary care institution. Commencing at 21 weeks of gestation, all antenatal care and follow-up visits are carried out at tertiary hospitals until delivery. High-risk pregnancies are always managed at tertiary hospitals.

**Table 2** Maternal and offspring characteristics at baseline and key information of REPRESENT database

Variables	Total number (No.)	Median (Q <sub>L</sub> –Q <sub>U</sub> ) <sup>a</sup>	Missing values (No., %)
Number of registrations per year	766 124	55 765 (35 020–96 973)	70 (0.01) <sup>c</sup>
Number of deliveries per year	758 147	73 904 (52 428–83 671)	150 (0.02) <sup>f</sup>
Maternal age (years)	764 069	27 (25–30)	2125 (0.27) <sup>c</sup>
Gestational week at registration (weeks)	748 851	13 (10–18)	17343 (2.26) <sup>c</sup>
Gestational week at delivery (weeks)	765 021	39.3 (38.4–40.1) <sup>d</sup>	725 (0.21) <sup>g</sup>
Gravidity (times) <sup>b</sup>	754 571	2 (1–3)	11 623 (1.52) <sup>c</sup>
Maternal height (m)	765 307	1.59 (1.56–1.62)	2416 (0.32) <sup>c</sup>
Pre-pregnancy body mass index (kg/m <sup>2</sup> )	702 512	20.29 (18.75–22.22)	64 923 (8.47) <sup>c</sup>
Birthweight (g)	761 962	3230 (2970–3500)	3784 (0.50) <sup>g</sup>
Length (cm)	765 515	50 (49–50)	231 (0.03) <sup>g</sup>
Education level (years)	715 680	N/A	50 228 (6.56) <sup>c</sup>
Illiterate and semi-illiterate		1718 (0.24) <sup>d</sup>	
1–6		17 887 (2.50) <sup>d</sup>	
7–9		211 201 (29.50) <sup>d</sup>	
10–12		157 612 (22.01) <sup>d</sup>	
>12		320 562 (44.77) <sup>d</sup>	
Unclear		6986 (0.98) <sup>d</sup>	
Parity ( <i>n</i> , %) <sup>c</sup>	709 817	0 (0–1)	56 377 (7.36) <sup>c</sup>
Nullipara		381 442 (53.74) <sup>d</sup>	
Multipara		328 375 (46.29) <sup>d</sup>	
Current smoking	765 909	N/A	285 (0.04) <sup>c</sup>
Yes		2424 (0.32) <sup>d</sup>	
No		763 485 (99.68) <sup>d</sup>	
Current alcohol consumption	765 909	N/A	285 (0.04) <sup>c</sup>
Yes		474 (0.06) <sup>d</sup>	
No		765 435 (99.94) <sup>d</sup>	
Mode of delivery	756 713	N/A	1524 (0.20) <sup>f</sup>
Vaginal delivery		510 891 (67.51) <sup>d</sup>	
Caesarean section		245 822 (32.49) <sup>d</sup>	
Multiple gestations	765 746	N/A	0 (0.00) <sup>f</sup>
Singleton		745 915 (97.41) <sup>d</sup>	
Multiple gestations		19 831 (2.59) <sup>d</sup>	

<sup>a</sup>Q<sub>L</sub>, lower quartile; Q<sub>U</sub>, upper quartile; N/A, not applicable.

<sup>b</sup>Gravidity refers to the number of pregnancies, including current and past pregnancies, no matter whether the pregnancies were interrupted or resulted in a live birth.

<sup>c</sup>Parity refers to the number of deliveries of a fetus at or after 24th gestational weeks, regardless of live birth or stillbirth. Nullipara indicates that a woman has no birth history, and multipara means a woman had more than one birth.

<sup>d</sup>No. (%).

<sup>e</sup>The number of pregnancies in antenatal care (766 194) is the denominator.

<sup>f</sup>The number of pregnancies in maternal delivery care (758 237) is the denominator.

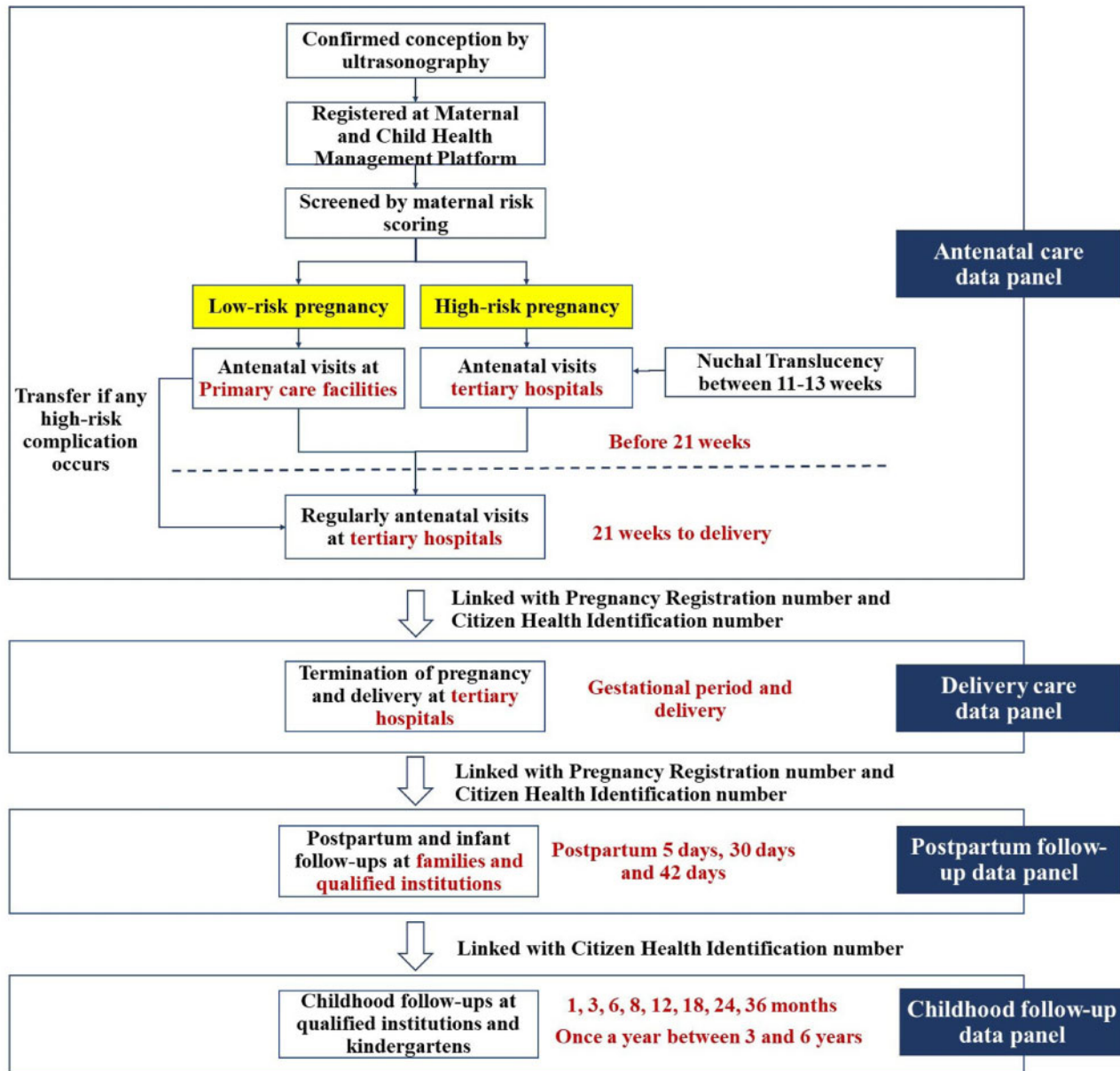
<sup>g</sup>The number of newborns in neonatal delivery care (765 746) is the denominator.

Typically, pregnant women undertake antenatal visits every 4 weeks during the first and second trimesters. From the third trimester (i.e. after 28 gestational weeks), the interval of antenatal visits is usually 1–2 weeks depending on the condition of pregnancy. As required by the Health Commission, information about all the visits are strictly documented by clinicians and stored in the Maternal and Child Health Management Platform in a timely manner.

After delivery, maternal family visits by primary health workers occur at 5 days and infant family visits at 30 days, as well as maternal examinations at 42 days at qualified

maternity departments. For women and infants with comorbidities, additional family visits may be conducted. The infant family visit at 30 days after delivery has been effective in Xiamen since 2010, whereas maternal family visits at 5 days after delivery began in 2014. The maternal examination at 42 days was initiated in 2013 on a voluntary basis.

Following the consistent and recommended follow-up workflow from conception to postpartum (Figure 3), the pregnancy registry was developed based on the Maternal and Child Health Management Platform. This platform has been effectively used for managing maternal and child



**Figure 3** Standard protocol of pregnancy care pattern in Xiamen

health across all healthcare institutions in Xiamen (accessed by attending clinicians through a fixed internal portal), which documents information and events about pregnant women from registration at their first trimester to postpartum, and includes childhood follow-up records. A birth certificate linking parental information is also automatically generated from this system. Attending clinicians report their assessment by completing specific forms of the platform, while pregnant women can also review their health reports through this system, as well as making appointments or booking a maternity bed.

Currently, the pregnancy registry consists of four data panels, namely, antenatal care, delivery care, postpartum

follow-up and childhood follow-up (Table 3). Of those, the antenatal care panel is concerned with registration visit, regular antenatal visits and antenatal screening. The delivery care panel comprises maternal and neonatal delivery records, where pregnancy termination outcomes and birth defects are separately documented. Any pregnancy loss, including abortion, induced labor, stillbirth/fetal death and the reasons for termination of pregnancy that occurred after the first antenatal visit, is recorded into a separate sheet by clinicians. The postpartum follow-up panel includes maternal family visits at 5 days after delivery and infant family visits at 30 days by primary care clinicians, as well as maternal examination at 42 days at

**Table 3** Data panels and key variables in the REPRESENT registry

Data panel	Subset	Categories	Key variables
Antenatal care	Registration visit	Socio-demographics	Maternal birth date, maternal age, education level, occupation, permanent residents, transient population, marital status, maternal age at marriage, parental age at present, current smoking and alcohol consumption
		Gestational characteristics	Gravidity, parity, date of the last menstrual period, number of vaginal deliveries, number of caesarean sections, and age at onset of menstruation
		Medical history	Hypertension, diabetes mellitus, history of surgical operation, thyroid dysfunction, and mental illness
		Maternal risk screening	Age (>35 or <18 years old), body mass index (>25 or <18.5 kg/m <sup>2</sup> ), height (<145 cm), use of assisted reproductive technology, history of abnormal pregnancy, diagnosed diseases at present (infectious diseases, respiratory diseases, cardiovascular diseases, haematological diseases, nervous system diseases, endocrinological diseases and immunological disease)
		Physical examinations	Height, weight, blood pressure, pre-pregnancy BMI, fundal height, and abdominal girth
		Medication use	Folic acid, contraceptives and other medications
		Gestational status	Gestational weeks, reactions of pregnancy, fetal movements, fetal heart and position of the fetus
		Laboratory tests	Haemoglobin count, numeration of leukocytes, blood platelet count, blood type, blood glucose, serum glutamic pyruvic transaminase, serum glutamic-oxaloacetic transaminase, total bilirubin, direct bilirubin, hepatitis B surface antigen, syphilis serodiagnosis, HIV, urine protein, urine ketone bodies, urine occult blood, serum creatinine and blood urea nitrogen
		Imaging examinations	B-mode ultrasound
		Gestational status	Gestational weeks, fetal movements, fetal heart, fetal position, and fetal presentation
		Physical examinations	Weight, blood pressure, fundal height and abdominal girth
		Laboratory tests	Oral glucose tolerance test, blood glucose, haemoglobin count, numeration of leukocytes, blood platelet count, blood type, serum glutamic pyruvic transaminase, serum glutamic-oxaloacetic transaminase, total bilirubin, direct bilirubin, hepatitis B surface antigen, syphilis serodiagnosis, HIV, urine protein, urine ketone bodies, urine occult blood, serum creatinine, blood urea nitrogen, triglyceride, total cholesterol, lactate dehydrogenase, creatine kinase, creatine kinase isoenzyme and prothrombin time
		Physical examinations	Height, weight
		Gestational information	Gestational weeks, date at the last menstrual period, crown-rump length, biparietal diameter, expected date of confinement, and multiple gestations
Delivery care	Maternal delivery care	History of birth defects	Maternal and familial history of birth defects (neural tube defects, Down's syndrome, Edward's syndrome)
		Serum biomarkers	Alpha fetoprotein, human chorionic gonadotropin, unconjugated oestriol, serum-free human chorionic gonadotropin, pregnancy-associated plasma protein, and nuchal translucency
		Pregnancy termination outcomes	Outcome of pregnancy (including delivery, abortion, induced labour, stillbirth/fetal death and others), date at outcome occurring, reason for termination of pregnancy
		Labor records	Timing of delivery, timing of uterine contraction, timing and methods of rupture of membranes, information about amniotic fluid, times of the first, second and third stages of labour, total stage of labour, mode of delivery, reason for dystocia, fetal position, placental disorders, length of umbilical cord, umbilical cord entangle neck, postpartum haemorrhage and blood pressure and delivery outcomes

(Continued)

Table 3 Continued

Data panel	Subset	Categories	Key variables
Postpartum follow-ups		Intrapartum medications	Amount and injection site of oxytocin and other medications
		Surgery records	Laceration perineum, laceration of vaginal douche and cervix, uterus ruptures, infection and other injuries
		Diagnoses	Gravidity, parity, number of fetus, gestational complications, gestational comorbidities and gestational hypertension
		Perinatal death	Stillbirth and death within 7 days after delivery
		Others	Nursing records
		Prenatal information	Gender, birth date and timing, gestational weeks, mode of delivery, birthweight, length and number of fetus
		Delivery descriptions	Premature rupture of membrane, placental abnormality, umbilical cord disorder, and Apgar scoring
		Prenatal death	Stillbirth and death within 7 days after delivery
		Others	Vaccination, breastfeeding, nursing records
		Maternal information	Age, race, educational level, gravity and parity
		Neonatal information	Birth date, gender, birthweight, gestational week at delivery, follow-up outcomes and use of induction
		Diagnosis of birth defect	Diagnosis date, diagnosis methods and type of defects
		Others	Exposure of medical diseases, medicines and other risk factors during the first trimester, the family history of birth defects
Childhood follow-ups		Maternal information	Blood pressure, breast condition, uterus condition and wound condition, follow-up outcomes
		Behaviour records	Personal hygiene, mentality, nutrition and feeding pattern
		Others	Referral and referral reason
		Infant information	Neonatal gender, neonatal birth date, gestational week at delivery, delivery pattern, neonatal asphyxia and Apgar score
		Parental information	Paternal occupation, paternal birth data, maternal occupation, maternal birth date and maternal gestational complications
		Physical examinations	Malformation, neonatal diseases screening (hearings, inherited metabolic diseases), weight, height, temperature, respiratory rate, complexion, icterus, bregma, eye, ear, nose, oral cavity, limbs, spine, cardiopulmonary auscultation, and abdominal palpation
		Daily life records	Feeding pattern, feeding volume, feeding times, emesis and excrement
		Others	Referral and referral reason
		Maternal information	Blood pressure, breast condition, lochia condition, uterus condition, wound condition and recovery levels
		Behaviour records	Feeding pattern and volumes
		Children information	Age and examination methods
		Physical examination	Weight, height, head width, chest width, abdomen width, bregma, temperature, respiratory rate, complexion, icterus, mentality, teeth number, caries number, eye, ear, nose, oral cavity, limbs, spine, heart, lungs, liver, spleen, hip joint, appetite and gait
		Mental examination	Physical development examination, nutrition examination, Ages and Stages Questionnaires estimation and neuropsychological development estimation
Daily life records	Feeding pattern, volume, and times		
Others	Disease diagnoses, referral, and referral reason		



maternity departments. The childhood follow-up panel is not part of the pregnancy care and is not conducted at maternity departments. However, all childhood data are collected through the Maternal and Child Health Management Platform and linked to the three other panels. In the first 3 years, the follow-ups are conducted at 1, 3, 6, 8, 12, 18, 24 and 36 months. After 36 months, the follow-ups are undertaken by primary care physicians annually at kindergartens until the child is 6 years of age.

### Data quality control

Since its inception, Xiamen Health Commission has implemented a rigorous approach to enforce data collection and quality control. During pregnancy, attending staff (clinicians, midwives and nurses) are required to file antenatal care visits through the Maternal and Child Health Management Platform. Information is manually entered by attending staff, except laboratory test data which are automatically uploaded. The original imaging records are saved locally, but the key diagnostic information is automatically uploaded to the platform.

After delivery, details on delivery care including maternal and neonatal records are completed electronically by attending clinicians. Any diagnosed birth defect must be filed separately. Postpartum follow-ups are completed according to the pre-specified protocol by attending clinicians or general practitioners (GPs).

### Linkage with other data platforms

The pregnancy registry offers a core data repository to construct a longitudinal follow-up of pregnancy care and outcomes from conception to postpartum and newborn tracking, as well as childhood follow-ups (up to 6 years). To develop REPRESENT, we linked the pregnancy registry with three other data platforms to increase data diversity and the length of follow-up. The current REPRESENT database has accumulated 11-year data (i.e. 2008–2019) and has the capacity to prolong follow-up with ongoing efforts of data accumulation from the four databases. As a result, one would be able to track pregnant women from pre-conception to a decade or longer after delivery, thus enabling long-term follow-up. Specifically, data are available both for exposures (e.g. historical diseases, medications, healthcare service and environmental factors) and long-term healthcare use and outcomes after delivery (Figure 4).

Similarly, the offspring could be followed up to adulthood by linking the databases, as each newborn is authorized a Citizen Health Information number at birth, which is used throughout his or her life. Thus, by linking databases, the health information of the offspring could be

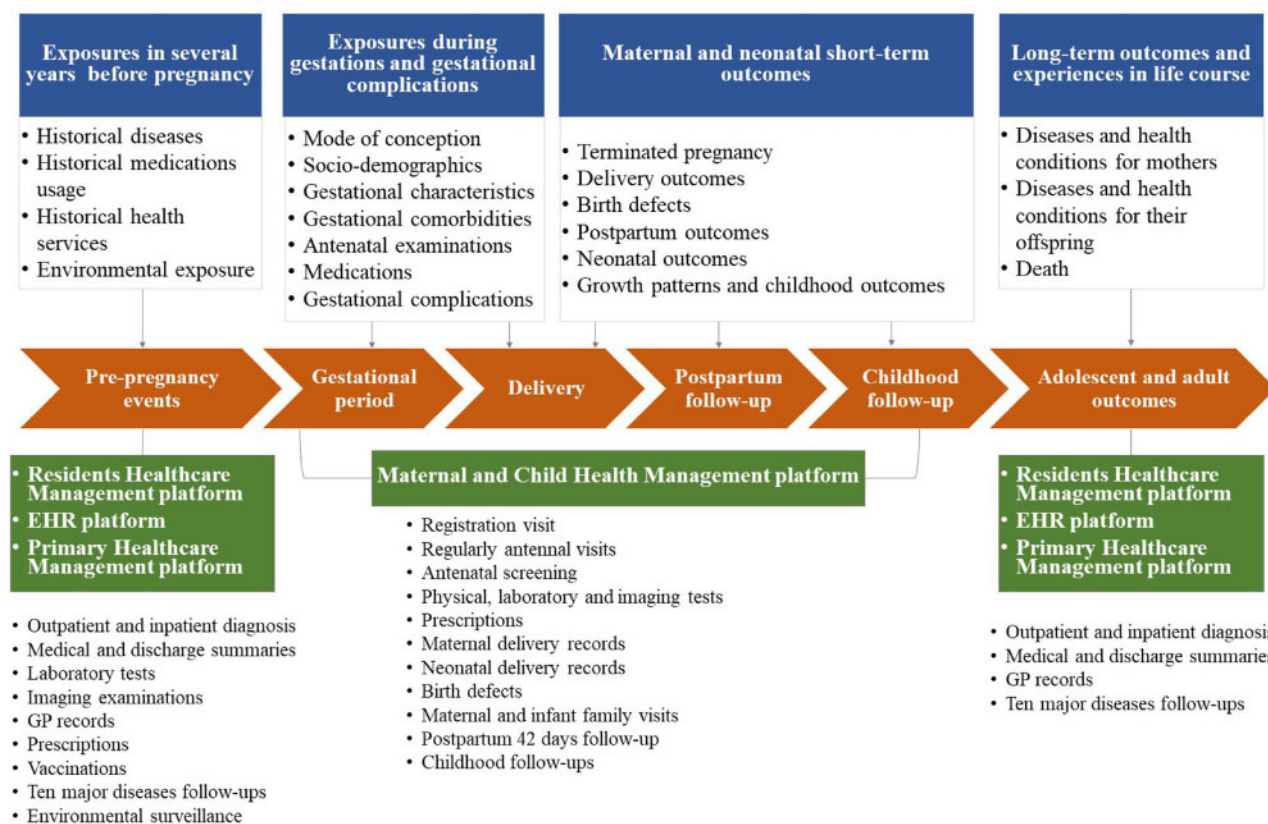
tracked over a life course (Figure 4). In addition, the offspring may be linked to their mothers by the family relationship record maintained in the Residents Healthcare Management Platform.

Several identification numbers are stored in the four platforms, including numbers for Citizen Health Identification, Pregnancy Registration, Birth Certificate number, and other internally used identification numbers across different institutions (e.g. medical institution number, inpatient number and outpatient number). We linked the pregnancy registry with the other platforms primarily by the Citizen Health Identification. In case of any missing identification number, approximate string matching is utilized based on other variables such as birth data, address, delivery date, medical institution number, inpatient number and outpatient number.

### Data extraction, validation, structuring and coding

As the initial step, experienced medical informatics experts, in consultation with the research team, extracted data from the Maternal and Child Health Management Platform to construct the fundamental database (i.e. prototype of the pregnancy registry). Because the Maternal and Child Health Management Platform underwent pilot testing in 2007 (when the unique identification number had not yet been adopted), the starting date for the pregnancy registry was January 2008. We randomly selected 1000 cases to validate the consistency of extracted data with raw data. After this assessment, the data extraction codes were adjusted, and iterative data extraction processes were performed until all extracted data were 100% complete and consistent for any given sample.

Most of the variables in the Maternal and Child Health Management Platform are designed as structured fields. Common classifications of gestational comorbidities, complications and diagnoses are listed in drop-down menus for selection by clinicians. Coding strategies for these data was implemented by experts in medical informatics. According to the Classification of Codes of Diseases released by Chinese National Standards Institute (GB/T14396-2016),<sup>12</sup> we retained the original classifications and encoded the drop-down menu by International Classification of Diseases 10th Revision (ICD-10). For complicated fields presented in semi-structured format, we refined the elements and categories existing in these fields, and coded the texts by matching the subject library of ICD-10 using the text clustering method. As a result, the disease categories and codes are standardized. We manually repeated checking the above process until the consistency rate was >95%. The platform adopted the National Health Standard Criteria to classify 23 main types of birth defects (WS 377.6-2013),<sup>13</sup> and



**Figure 4** Linking pregnancy registry with other data platforms; EHR, Electronic Healthcare Records

Guidelines for reporting complicated and micro malformation in China Maternal and Child Health Surveillance,<sup>14</sup> China 2019 Congenital Structural Defects Program<sup>15</sup> and China 2017 Disease list of Congenital Metabolic Defects Program,<sup>16</sup> were used for identify other birth defects. In addition, medications were encoded using the National Drugs Coding Criteria (PYID\_5.0).<sup>17</sup>

Although the data linkage has been completed and successful, the work is still ongoing to transform the semi-structured EHR data into more structured records. In particular, medication history, inpatient diagnoses, surgical records and inpatient medical services before and after pregnancy require further refinements. Ultimately, a persistent, structured population-based healthcare database capturing information about pregnant women and their offspring will be established.

### Patient privacy and ethics review

Informed consent is a compulsory process for all the residents in Xiamen who are registered with a Citizen Health Identification number (not just pregnant women). At the registration, the residents or their guardians (typically children's parents) should sign informed consent to authorize the development of personal life-course electronic

health files. All medical and healthcare records, which constitute the personal life-course electronic health files, are stored in the four healthcare management platforms. The residents can also review the electronic health file online and revise the authorization about the degree of openness and permissions.

To protect patient privacy and to ensure data security, the process of data access, data extraction, data processing and data analysis was accomplished through a specially allocated server at Xiamen Health and Medical Big Data Center (which is not externally accessible). Experienced software engineers, who manage the platforms, undertake the primary responsibility for data extraction, whereas research investigators cannot access identification and sensitive patient information without permission from the data manager. All investigators sign a data confidentiality agreement before being given data access. The present study was approved by West China Hospital's Ethics Committee (2019-825) and registered at *clinicaltrials.gov* (NCT04222621).

### Data resource use

We have identified a number of potential uses of the registry. Firstly, REPRESENT has the capacity to address

important research questions by linking the four major healthcare data platforms. For instance, one would be able to determine whether pre-pregnancy exposures (e.g. medications or vaccines) can impact gestational outcomes (e.g. hypertensive diseases of pregnancy),<sup>18</sup> birth outcomes such as birth defects,<sup>19,20</sup> maternal outcomes (e.g. postpartum hemorrhage)<sup>21</sup> and even long-term childhood developments (e.g. attention deficit hyperactivity disorder).<sup>18,22,23</sup> Similarly, it is feasible to investigate the effect of fetus exposure to subsequent diseases (e.g. childhood cancer)<sup>24,25</sup> and the growth pattern in childhood and adolescence. A new set of evidence may then be generated to improve management of pregnant women and their offspring.

Secondly, due to the large population and the high level of details documented in REPRESENT, there are great opportunities to investigate rare conditions. For instance, given the high caesarean section rate in the past decades, reducing serious but rare adverse outcomes (e.g. uterus ruptures)<sup>26</sup> becomes an imperative. With sufficient numbers of events and adequate information (e.g. antenatal examinations, nursing records), a prediction model may be formulated for uterus ruptures, which could enable early risk assessment and timely interventions. In addition, with a substantial volume of variables and individuals in REPRESENT, it is feasible to ascertain causation, which is another unique feature of this registry. One such example is a classic regression discontinuity design in which a small sample is selected from a very narrow interval around 1500 g as a threshold, to estimate the causal effect of intensified medical treatment on 1-year mortality among very low-birth-weight babies (weighing <1500 g).<sup>27</sup> A large number of newborns with birth records and a rich set of covariates are required for such investigation.

Thirdly, REPRESENT is built on a specific region in China whereby a tiered healthcare system (i.e. primary-secondary-tertiary care) has been well developed. Findings from such populations are pertinent for enhancing healthcare policy in China. In particular, one may assess the influence of behaviors and incentives of the childbearing population, as a result of the two-child policy and changes in socio-economic status.<sup>28,29</sup> Therefore, it has the capacity to assess the operational efficiency of the tiered healthcare system, and develop policy recommendations regarding resource allocation and referral pathways.

Fourthly, REPRESENT has documented a high level of details about healthcare conditions and resources use during the life course in a population-level setting. Causations for adverse pregnancy outcomes are often not established due to data inadequacy, especially when assessing the effects of healthcare services. An example is the apparent association between peri-conceptual or pregnancy exposure to human papillomavirus (HPV) vaccination and the

risk of adverse perinatal outcomes.<sup>30</sup> Historically it is difficult to acquire accurate records of HPV vaccination and perinatal outcomes, such as spontaneous abortion and birth defects. In contrast, our new database offers a unique opportunity to bridge this methodological gap and the findings may be potentially generalizable to other pregnancy populations in China.

Finally, REPRESENT permits comparisons about the care pattern and outcomes among populations in China and other countries.<sup>31,32</sup> In summary, given the large volume of newborns each year in China, we believe that the resulting evidence would have important implications for changing practice and policy.

## Strengths and weaknesses

REPRESENT has several advantages. First of all, it is the first population-based, long-term follow-up registry of pregnant women and their offspring in China. By linking four major healthcare data platforms, one can capture information on a diversity of exposures before pregnancy as well as long-term outcomes for both mothers and infants, in addition to tracking data during gestation and delivery, which provides the chance to explore the causality between exposure before and during pregnancy (e.g. medication usage) and short- and long-term outcomes.

Second, one may access a large and comprehensive list of healthcare information spanning from gestation to postpartum and childhood follow-ups, which not only covers pregnancy information at each maternity department in Xiamen, but also includes a complete set of information about diagnoses, descriptions, medical histories, medication usage, physical examinations, laboratory tests and imaging results documented according to consistent workflow and formats. The data volume is substantial with over 800 variables being documented. In the past 11 years (January 2008 to March 2019) REPRESENT has accumulated data concerning 766 194 pregnancies and 765 746 newborns.

Third, the quality of data is assured because of the policy for unified data collection implemented by the Health Commission, particularly data collected via the Maternal and Child Health Management Platform, which represents the foundation for the pregnancy registry. Furthermore, medical history, gestational history, medication usage, family history and family relationships can be exactly extracted and analyzed from four data platforms, which avoids possible recall bias from the common Pregnancy Registry.

It also has several potential weaknesses. A limitation of REPRESENT is the lack of national representation, especially in view of the socio-economic and cultural disparities

between eastern and western China, and the epidemiological characteristics of the Xiamen population may differ from other regions. To overcome this shortcoming, it is our intention to develop a common data platform that links databases across regions in China by forging a collaborative network of partnership institutions in the future. Secondly, live birth bias may be present in our database, as certain early pregnancy losses (e.g. pregnancy loss before clinical pregnancy and early miscarriage before registration) are less likely to be recorded. In our database, a separate sheet about pregnancy termination outcomes was available, which can document any pregnancy loss by clinicians, including abortion, induced labor, stillbirth/fetal death and the reasons for termination of pregnancy, occurring after the first antenatal visit. Adjusting the risk factors that both influence pregnancy losses and the outcomes of interest during the statistical analysis may mitigate the magnitude of live birth bias. In addition, other forms of selection biases, such as survivor bias and indication bias, may be present when using the database. However, given its nature as a population-based database and the comprehensiveness of variables documented, this database may offer the best available data source for this population in China.

### Data resource access

The Chinese Evidence-based Medicine Center, affiliated with West China Hospital, Sichuan University, is responsible for developing REPRESENT. Both the Chinese Evidence-based Medicine Center and Xiamen Health and Medical Big Data Center are jointly responsible for granting the use of the registry data. Research institutions can apply to access the registry data by submitting a formal study protocol, subject to approval by the Xiamen Health and Medical Big Data Center and the Chinese Evidence-based Medicine Center. Ethical review and research registration are mandatory for all studies. Research institutions can apply for access to the registry data by submitting a formal study protocol, subject to approval by Xiamen Health and Medical Big Data Center and Chinese Evidence-based Medicine Center ([http://www.xmyys.com.cn/news\\_info\\_204.html](http://www.xmyys.com.cn/news_info_204.html)).

### Supplementary data

Supplementary data are available at *IJE* online.

### Funding

The study was supported by funding from the National Natural Science Foundation of China (71974138, 71704122), National Key

Research and Development Program of Reproductive Health & Major Birth Defects Control and Prevention (2016YFC1000406), and National Science and Technology Major Project (2018ZX10302206-004-005).

### Acknowledgements

The authors would like to acknowledge the following individuals and institutions who have made substantial contributions to the establishment of REPRESENT: Xiamen Health and Medical Big Data Center (Shaomei Chen, Huimei Li, Ling Li), College of Public Administration, Huazhong University of Science & Technology (Jindong Ma, Yang Song), Department of Obstetrics, Xiamen Maternal and Child Care Hospital (Xueqin Zhang, Quanfeng Wu). X.S. and J.T. are guarantors.

### Conflict of interest

None declared.

### References

1. National Health Committee of People's Republic of China. *China Health Statistics Yearbook*. Beijing: Peking Union Medical College Press, 2018.
2. Guo Y, Zakus D, Liang H. China: policy and practice of MCH since the early 1990s. *Matern Child Health J* 2008;12:139–48.
3. Liang J, Mu Y, Li X *et al*. Relaxation of the one child policy and trends in caesarean section rates and birth outcomes in China between 2012 and 2016: observational study of nearly seven million health facility births. *BMJ* 2018;360:k817.
4. Feng XL, Guo S, Hipgrave D *et al*. China's facility-based birth strategy and neonatal mortality: a population-based epidemiological study. *Lancet* 2011;378: 1493–500.
5. The Central Committee of the Communist Party of China and State Council. The Central Committee of the Communist Party of China and the State Council issued the Outline of Healthy China 2030. 2016. <http://www.nhc.gov.cn/guihuaxxs/s3586s/201610/21d120c917284007ad9c7aa8e9634bb4.shtml> (20 October 2019, date last accessed).
6. Zeng Y, Hesketh T. The effects of China's universal two-child policy. *Lancet* 2016;388: 1930–938.
7. Ministry of Health of People's Republic of China. Technical solution for building hospital information platform based on electronic medical record (version 1.0). 2011. <http://www.nhc.gov.cn/mohwsbwstjxxzx/s7968/201103/51079.shtml> (18 October 2019, date last accessed).
8. Zhao J, Zhang Z, Guo H *et al*. E-health in China: challenges, initial directions, and experience. *Telemed J E Health* 2010;16: 344–49.
9. Yi M, Li D, Jun Z, Qi L, Xia M, Zheng L. Evaluation of application effect of system for Chinese maternal and child health surveillance. *Matern Child Health Care China* 2011;26: 5477–480.
10. Ministry of Health of People's Republic of China. Administrative Regulation on Health Care during Pregnancy and Childbirth and Good Practice for Health Care during Pregnancy and Childbirth. <http://www.nhc.gov.cn/fys/s3581/201107/8d09ba60c19545e3b80fa65328183537.shtml> (26 April 2020, date last accessed).

11. National Health and Family Planning Commission of People's Republic of China. Norms for risk assessment and maternal management of pregnancy. 2017. <http://www.nhc.gov.cn/fys/gzbs/201711/9c3dc9b4a8494d9a94c02f890e5085b1.shtml> (3 December 2019, date last accessed).
12. National Health and Family Planning Commission of People's Republic of China and China. National Standardization Management Committee. Classification and codes of diseases (GB/T 14396-2016), 2016. <http://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=8127A7785CA677952F9DA062463CBC41> (30 October 2019, date last accessed).
13. National Health and Family Planning Commission of People's Republic of China. Women's health basic data set - part 6: birth defect surveillance, 2013. <http://www.nhc.gov.cn/ewebeditor/uploadfile/2018/09/20180921141035465.pdf> (30 October 2019, date last accessed).
14. China Maternal and Child Health Surveillance. China birth defects surveillance system: Guidelines for reporting complicated and micro malformations, 2012. <http://www.mchscn.org/admin/wenjian/wxjxbgbz-2012.pdf> (30 October 2019, date last accessed).
15. National Health Committee of People's Republic of China. 1 minute to understand the Congenital Structural Defects Program. 2019. <http://www.nhc.gov.cn/fys/s3590/201903/9e08496ee740468a93764c4afd3ff4d8.shtml> (30 October 2019, date last accessed).
16. March of Dimes Birth Defects Foundation of China. Disease list of congenital metabolic defects program, 2017. <http://www.csqx.org.cn/content.aspx?id=381707814440> (30 October 2019, date last accessed).
17. National Health Committee of People's Republic of China. China Drug Supply Information Platform, 2018. <http://cdisip.nhc.gov.cn/PriceBase/YPIDList.aspx?key=yxid&cvalue=> (30 October 2019, date last accessed).
18. Eilertsen EM, Gjerde LC, Reichborn-Kjennerud T *et al*. Maternal alcohol use during pregnancy and offspring attention-deficit hyperactivity disorder (ADHD): a prospective sibling control study. *Int J Epidemiol* 2017;46:1633–640.
19. Pasternak B, Hviid A. Use of proton-pump inhibitors in early pregnancy and the risk of birth defects. *N Engl J Med* 2010;363:2114–123.
20. Interrante JD, Ailes EC, Lind JN *et al*. Risk comparison for prenatal use of analgesics and selected birth defects, National Birth Defects Prevention Study 1997–2011. *Ann Epidemiol* 2017;27:645–53.e2.
21. Hanley GE, Smolina K, Mintzes B, Oberlander TF, Morgan SG. Postpartum hemorrhage and use of serotonin reuptake inhibitor antidepressants in pregnancy. *Obstet Gynecol* 2016;127:553–61.
22. Morales DR, Slattery J, Evans S, Kurz X. Antidepressant use during pregnancy and risk of autism spectrum disorder and attention deficit hyperactivity disorder: systematic review of observational studies and methodological considerations. *BMC Med* 2018;16:6.
23. Ystrom E, Gustavson K, Brandlistuen RE *et al*. Prenatal exposure to acetaminophen and risk of ADHD. *Pediatrics* 2017;140:e20163840.
24. Hargreave M, Jensen A, Hansen MK *et al*. Association between fertility treatment and cancer risk in children. *JAMA* 2019;322:2203–210.
25. Man KKC, Chan EW, Ip P *et al*. Prenatal antidepressant use and risk of attention-deficit/hyperactivity disorder in offspring: population based cohort study. *BMJ* 2017;357:j2350.
26. Vandenberghe G, De Blaere M, Van Leeuw V *et al*. Nationwide population-based cohort study of uterine rupture in Belgium: results from the Belgian Obstetric Surveillance System. *BMJ Open* 2016;6:e010415.
27. Almond D, Doyle JJ, Kowalski AE, Williams H. Estimating marginal returns to medical care: evidence from at-risk newborns. *Q J Econ* 2010;125:591e634.
28. Liao Z, Zhou Y, Li H, Wang C, Chen D, Liu J. The rates and medical necessity of cesarean delivery in the era of the two-child policy in Hubei and Gansu Provinces, China. *Am J Public Health* 2019;109:476–82.
29. Li H-T, Xue M, Hellerstein S *et al*. Association of China's universal two child policy with changes in births and birth related health factors: national, descriptive comparative study. *BMJ* 2019;366:l4680.
30. Tan J, Xiong Y-Q, He Q *et al*. Peri-conceptual or pregnancy exposure of HPV vaccination and the risk of spontaneous abortion: a systematic review and meta-analysis. *BMC Pregnancy Childbirth* 2019;19:302.
31. Fok D, Aris IM, Ho J *et al*. A comparison of practices during the confinement period among Chinese, Malay, and Indian Mothers in Singapore. *Birth* 2016;43:247–54.
32. Tan J, Ren Y, Qi Y *et al*. The pattern of gestational weight gains among Chinese women: a repeated measure analysis. *Sci Rep* 2018;8:15865.