

# BMJ Open Maternity Log study: a longitudinal lifelog monitoring and multiomics analysis for the early prediction of complicated pregnancy

Junichi Sugawara,<sup>1,2</sup> Daisuke Ochi,<sup>1,3</sup> Riu Yamashita,<sup>1</sup> Takafumi Yamauchi,<sup>1,3</sup> Daisuke Saigusa,<sup>1</sup> Maiko Wagata,<sup>1,2</sup> Taku Obara,<sup>1</sup> Mami Ishikuro,<sup>1</sup> Yoshiki Tsunemoto,<sup>3</sup> Yuki Harada,<sup>1</sup> Tomoko Shibata,<sup>1</sup> Takahiro Mimori,<sup>1</sup> Junko Kawashima,<sup>1</sup> Fumiki Katsuoka,<sup>1</sup> Takako Igarashi-Takai,<sup>1</sup> Soichi Ogishima,<sup>1</sup> Hirohito Metoki,<sup>4</sup> Hiroaki Hashizume,<sup>1</sup> Nobuo Fuse,<sup>1,2</sup> Naoko Minegishi,<sup>1</sup> Seizo Koshiba,<sup>1</sup> Osamu Tanabe,<sup>1,5</sup> Shinichi Kuriyama,<sup>1,2</sup> Kengo Kinoshita,<sup>1</sup> Shigeo Kure,<sup>1,2</sup> Nobuo Yaegashi,<sup>1,6</sup> Masayuki Yamamoto,<sup>1,2</sup> Satoshi Hiyama,<sup>1,3</sup> Masao Nagasaki<sup>1,2</sup>

**To cite:** Sugawara J, Ochi D, Yamashita R, *et al.* Maternity Log study: a longitudinal lifelog monitoring and multiomics analysis for the early prediction of complicated pregnancy. *BMJ Open* 2019;**9**:e025939. doi:10.1136/bmjopen-2018-025939

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-025939>).

Received 11 August 2018  
Revised 20 December 2018  
Accepted 27 December 2018



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Junichi Sugawara;  
[jsugawara@med.tohoku.ac.jp](mailto:jsugawara@med.tohoku.ac.jp)

## ABSTRACT

**Purpose** A prospective cohort study for pregnant women, the Maternity Log study, was designed to construct a time-course high-resolution reference catalogue of bioinformatic data in pregnancy and explore the associations between genomic and environmental factors and the onset of pregnancy complications, such as hypertensive disorders of pregnancy, gestational diabetes mellitus and preterm labour, using continuous lifestyle monitoring combined with multiomics data on the genome, transcriptome, proteome, metabolome and microbiome.

**Participants** Pregnant women were recruited at the timing of first routine antenatal visits at Tohoku University Hospital, Sendai, Japan, between September 2015 and November 2016. Of the eligible women who were invited, 65.4% agreed to participate, and a total of 302 women were enrolled. The inclusion criteria were age  $\geq 20$  years and the ability to access the internet using a smartphone in the Japanese language.

**Findings to date** Study participants uploaded daily general health information including quality of sleep, condition of bowel movements and the presence of nausea, pain and uterine contractions. Participants also collected physiological data, such as body weight, blood pressure, heart rate and body temperature, using multiple home healthcare devices. The mean upload rate for each lifelog item was ranging from 67.4% (fetal movement) to 85.3% (physical activity), and the total number of data points was over 6 million. Biospecimens, including maternal plasma, serum, urine, saliva, dental plaque and cord blood, were collected for multiomics analysis.

**Future plans** Lifelog and multiomics data will be used to construct a time-course high-resolution reference catalogue of pregnancy. The reference catalogue will allow us to discover relationships among multidimensional phenotypes and novel risk markers in pregnancy for the future personalised early prediction of pregnancy complications.

## Strengths and limitations of this study

- This is the first study designed to collect longitudinal lifelog information through healthcare devices, self-administered questionnaires using smartphones and varieties of biospecimens throughout pregnancy.
- Longitudinal, continuous, individual lifelog data with a high acquisition rate will enable us to assess dynamic physiological changes throughout pregnancy.
- Multiomics data will make it possible to understand the complex mechanisms of multifactorial pregnancy-related diseases.
- Potential limitations are the limited sample size and participant recruitment only at a tertiary hospital for high-risk populations.
- Inclusion criteria of the present study limited the eligibility to pregnant women with age  $>20$  years and the ability to access the internet using a smartphone.

## INTRODUCTION

The incidence of pregnancy-related disorders, including hypertensive disorders of pregnancy (HDP), gestational diabetes mellitus (GDM) and preterm delivery has been increasing worldwide.<sup>1-4</sup> These multifactorial conditions are caused by an interaction of genetic factors and environmental factors.<sup>5,6</sup> Recent reports suggest that continuous lifestyle monitoring using wearable biosensors provides important information on latent physiological changes that are exhibited prior to the onset of disease.<sup>7</sup> Using these monitors, environmental factors may be estimated more accurately than by using conventional questionnaires.

For these reasons, we have designed a prospective cohort study for pregnant women, the Maternity Log study (MLOG). In this study, pregnant women upload daily information and physiological data using multiple home healthcare devices. In addition, a variety of biospecimens are collected for multiomics analysis.

To the best of our knowledge, this study will be the first to integrate multiomics analyses and objective data on environmental factors, including daily lifelog data, in pregnant women. This study may demonstrate correlations between specific lifelog patterns and pregnancy-related physiological changes, such as blood pressure, gestational weight gain and onset of obstetric diseases. Furthermore, studies on associations among lifelog patterns, plasma and urine metabolomes, transcriptomes and genomic variations may reveal relationships among multidimensional phenotypes and lead to identification of novel risk markers in pregnancy for the future personalised early prediction of pregnancy complications, for example, HDP, gestational diabetes and preterm labour.

## COHORT DESCRIPTION

### Study setting

The aim of the MLOG study is to construct a time-course high-resolution reference catalogue of bioinformatic data in pregnancy and thereby develop methods for early prediction of obstetric complications, through integrated analysis of daily lifelogs and multiomics data, *that is*, maternal genomes, transcriptomes, metabolomes and oral microbiomes.

The MLOG study is a prospective, add-on cohort study, built on a birth-generation and three-generation cohort study established by the Tohoku Medical Megabank Organization (ToMMo) (TMM BirThree Cohort Study)<sup>8</sup> in order to elucidate the mechanisms of complicated multifactorial diseases in mothers and children in the wake of the Great East Japan Earthquake in 2011. Epidemiological data from extensive questionnaire surveys and accurate clinical records, including birth outcomes, can be abstracted from the integrated biobank of the ToMMo.<sup>8</sup> TMM BirThree Cohort Study was started in July 2013 in one obstetric clinic and expanded throughout Miyagi Prefecture, and approximately 50 obstetric clinics and hospitals (including Tohoku University Hospital) participated in the recruiting process. We planned to recruit 20 000 pregnant women as probands, and her family members from three generations, which is a total of over 70 000 participants.<sup>8</sup> Written informed consent was obtained from all participants by the genome medical research coordinators (GMRCs).

### Patient and public involvement

Patients or the public were not directly involved in the development of the research question or the design of the study. The main results will be made available in the public domain.

## Participants

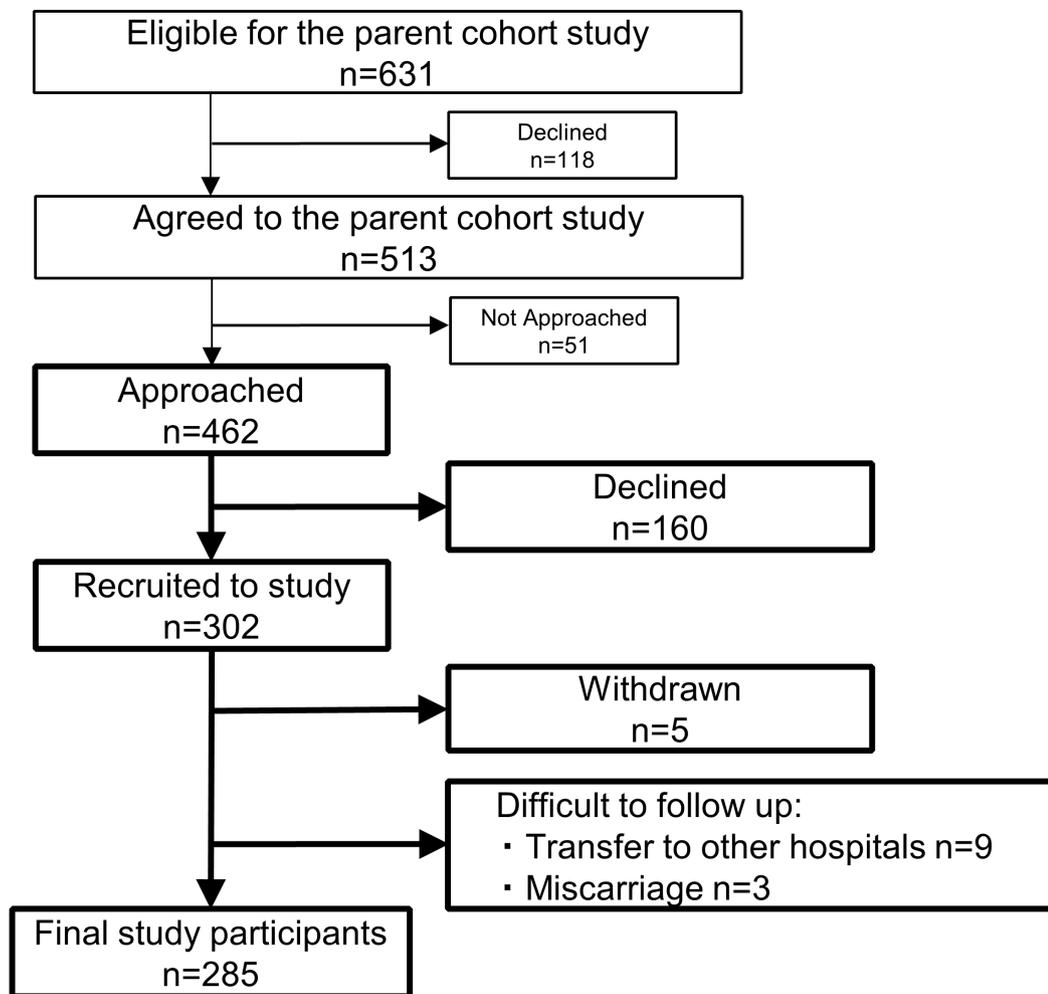
Participants were recruited at a first routine antenatal visit at Tohoku University Hospital, Sendai, Japan, between September 2015 and November 2016. A flow chart of the recruitment process is shown in [figure 1](#). GMRCs at Tohoku University Hospital approached eligible pregnant women for TMM BirThree Cohort Study (n=631), and patients who already agreed to participate in TMM BirThree Cohort Study (n=513) were assessed for eligibility for the MLOG study. Finally, 462 pregnant women were asked to provide informed consent for the MLOG study. A total of 302 women were enrolled. The inclusion criteria were the age  $\geq 20$  years and the ability to access the internet using a smartphone in the Japanese language. Participants were excluded after enrolment if termination of pregnancy, abortion or transfer to another institution for emergency care occurred before delivery, or if they withdrew consent for any reason.

## Outline of study protocol

The study protocol consisted of blood and urine sampling, saliva and dental plaque sampling, self-administered daily lifelog data collection and data upload from multiple healthcare devices through a smartphone. An overview of the protocol is provided in [figure 2](#). In Japan, routine antenatal visits, including ultrasounds, are scheduled every 4 weeks from early pregnancy (<12 weeks) to 23 weeks of gestation, every 2 weeks from 24 weeks to 35 weeks and every week from 36 weeks to delivery.<sup>9</sup> Lifelog data collection was continued throughout pregnancy and until 1 month after delivery. Optional data collection could be continued up to 180 days after delivery.

## Blood and urine sampling

Blood samples were collected three times from each participant; the first sample was collected between 12 weeks and 24 weeks of gestation, the second between 24 weeks and 36 weeks, and the third at 1 month after delivery. A maximum of 13 mL of blood was collected each time, from which serum and plasma were separated to be stored at  $-80^{\circ}\text{C}$  until the time of analysis. An aliquot of blood (2.5 mL) was stored in a PAXgene tube (Becton, Dickinson and Company, Franklin Lakes, New Jersey, USA) at  $-80^{\circ}\text{C}$  until the time of RNA extraction for transcriptome analysis. Genomic DNA was extracted from mononuclear cells using an Autopure extractor (Qiagen, Venlo, The Netherlands). Approximately 10 mL of cord blood was collected from the umbilical vein in a PAXgene tube for storage at  $-80^{\circ}\text{C}$  and in an EDTA 2K tube (Becton, Dickinson and Company) for separation of plasma to be stored at  $-80^{\circ}\text{C}$ . Urine samples (10 mL) were collected at each antenatal visit; when participants were admitted to the hospital ward, urine was collected once weekly. Urine samples were immediately transferred and stored at  $-80^{\circ}\text{C}$  until the time of analysis.



**Figure 1** Flow chart of Maternity Log (MLOG) study (MLOG) participants.

### Saliva and dental plaque sampling

Samples of saliva and dental plaque were collected three times from each participant, at the same time points as blood collection. Approximately 3 mL of saliva was collected using a 50 mL conical centrifuge tube (Corning, Inc, Corning, New York, USA) and stored at  $-80^{\circ}\text{C}$  until analysis. Dental plaque was sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM EDTA; pH, 8.0) and immediately stored at  $-80^{\circ}\text{C}$  until the time of sample processing.

### Lifelog data collection

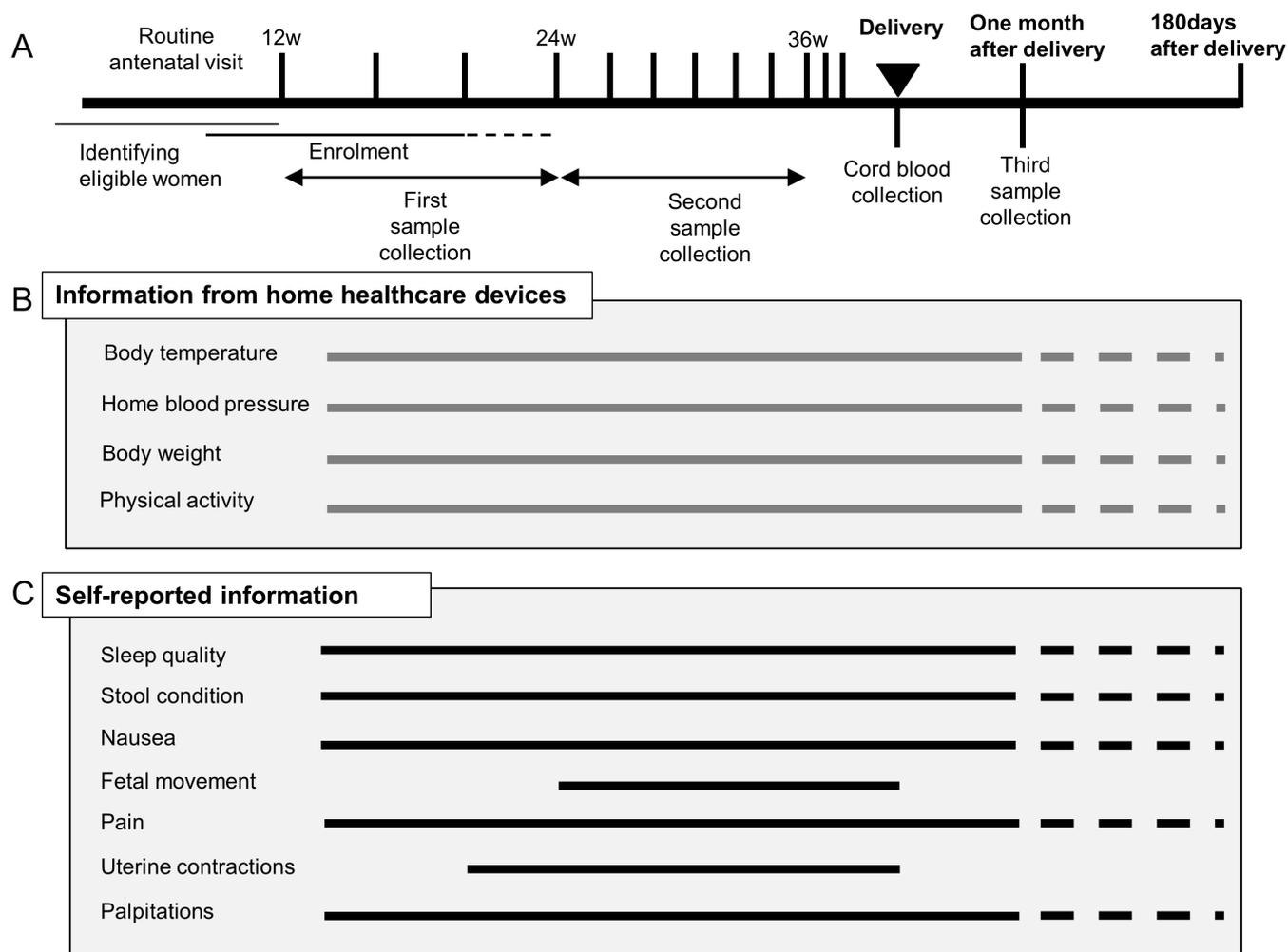
Based on previous publications on the utility for risk assessment of pregnancy-related diseases, we selected several lifelog parameters to employ in this study, that is, body temperature,<sup>10</sup> home blood pressure,<sup>11</sup> body weight<sup>12</sup> and physical activity (calorie expenditure),<sup>13</sup> as well as self-administered information such as sleep quality,<sup>14</sup> condition of stool,<sup>15</sup> severity of nausea,<sup>16</sup> fetal movement,<sup>17</sup> severity of pain,<sup>18</sup> uterine contractions<sup>19</sup> and palpitations.<sup>20</sup> Body temperature, home blood pressure, body weight and physical activity were uploaded from multiple healthcare devices through a smartphone. The self-administered

information described above was input manually on mobile applications created for this study.

Data collection was started after obtaining informed consent and after giving detailed instructions for the use of the healthcare devices. These applications tracked quality of sleep; condition of stool using the Bristol Scale<sup>21–23</sup>; severity of nausea using the Pregnancy-Unique Quantification of Emesis and nausea (PUQE) score<sup>24,25</sup>; headache, toothache, lumbago and upper and lower abdominal pain using a numerical rating scale (NRS) score; the number of perceived uterine contractions; palpitations; and fetal movement using a modified count-to-10 fetal movement chart.<sup>26,27</sup>

Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction (ranked from satisfied to poor using a numeric scale of 0–4) and the number of nocturnal awakenings (0–6).

The Bristol stool form scale was originally developed to assess constipation and diarrhoea,<sup>21,22</sup> and its use has been spread widely to evaluate functional bowel disorders.<sup>22</sup> Using the Bristol scale, stool is classified into seven types according to cohesion and surface cracking.<sup>21,22</sup>



**Figure 2** Overview of the MLOG study protocol. (A) Participant timeline for the MLOG study. (B) Physiological information collected using healthcare devices. Specific measures were uploaded each day from the time of enrolment (solid horizontal lines). Participants had the option to continue uploading data until 180 days after delivery (dashed horizontal lines). (C) Daily lifelogs of self-reported information using a smartphone application. Basic lifelog information was input manually from the time of enrolment (solid horizontal lines). Participants had the option to continue uploading data until 180 days after delivery (dashed horizontal lines). Fetal movement and uterine contractions were recorded from 24 weeks and 20 weeks of gestation, respectively.

The PUQE score<sup>24 25</sup> was developed to estimate the severity of nausea and vomiting in pregnancy and quantifies the number of daily vomiting and retching episodes and the length of nausea in hours (over the preceding 12 hours). The total score ranges from 3 (no symptoms) to 15, and higher scores are correlated with increasing severity of nausea and vomiting.<sup>24 25</sup>

In the NRS score for headache, toothache, lumbago and upper and lower abdominal pain, the total score ranges from 0 (no pain) to 10 (maximum ever experienced).

Uterine contractions and palpitations were evaluated using definitions determined for the current study. Uterine contractions were assessed using the number of perceived contractions per day, ranging from 0 to more than 5. The count-to-10 method was originally developed to assess fetal well-being by recording the time, in minutes, required to count 10 fetal movements.<sup>26</sup> More recently, a modified count-to-10 method has been proposed:

pregnant women are advised to start counting when they feel the first movement, then record the time required to perceive an additional nine movements.<sup>27</sup> Pregnant women are encouraged to select a 2-hour period when they feel active fetal movements and are instructed to count kicking and rolling movements in a favourable maternal position after 24 weeks of gestation.

The applications also collected dietary logs and the medications taken on the day before and the day of the antenatal visit on which blood or urine samples were collected.

Daily home blood pressure, body weight, body temperature and physical activity were measured as described below with home healthcare devices and uploaded through wireless communications using mobile applications on a smartphone. Daily home blood pressure was measured twice daily using an HEM-7510 monitor (OMRON Healthcare, Kyoto, Japan): within 1 hour of

awakening in the morning and just before going to bed at night. Body weight was measured using an HBF-254C metre (OMRON Healthcare) once daily within 1 hour of awakening in the morning. Daily body temperature was evaluated using an MC-652LC digital thermometer (MC-652LC; OMRON Healthcare) just after awakening. Physical activity was assessed using an HJA-403C pedometer (HJA-403C; OMRON Healthcare) to count steps and calculate calorie expenditure.

### Clinical and epidemiological information

Baseline clinical information and maternal and neonatal outcomes (eg, maternal age, clinical data and findings from each antenatal visit, gestational age at delivery, type of delivery, birth weight and maternal and fetal complications) were obtained from the medical records of Tohoku University Hospital. Epidemiological data, including extensive questionnaire surveys by TMM BirThree Cohort Study, can be obtained from the ToMMo integrated biobank.<sup>8</sup>

### Database

A customised laboratory information management system (LIMS) was established to track all biospecimens. All data were transferred to the TMM integrated database after two-step anonymisation in a linkable fashion.

Data handling was strictly regulated under Health Insurance Portability and Accountability Act of 1996 US Security and Privacy Rules<sup>28 29</sup> and the Act on the Protection of Personal Information.<sup>30</sup> Security control at our facility has been described previously.<sup>31</sup>

### Omics analysis

#### Whole-genome sequencing

To minimise amplification bias, we adopted a PCR-free library preparation method. After performing library quality control (QC) using the quantitative MiSeq method,<sup>32</sup> libraries were sequenced on HiSeq 2500 Sequencing System (Illumina, San Diego, California, USA) to generate 259 bp, paired-end reads. We generated the sequencing data at over 12.5× coverage on average, and we identified variants using the alignment tool BWA-MEM (V.0.7.5a-r405) with the default option. Single nucleotide variants (SNVs) and indels were jointly called across all samples using Genome Analysis Tool Kit's HaplotypeCaller (V.8). Default filters were applied to SNV and indel calls using the GATK's Variant Quality Score Recalibration approach. The human reference genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC\_007605 (Human Gamma Herpesvirus 4). The complete fasta file named hg19\_tommo\_v2.fa is available from iJGVD website (<http://ijgvd.mega-bank.tohoku.ac.jp>).<sup>33</sup> For the quality assurance, we have checked the ratio of the bases with the phred quality score over 30, the total variant numbers in each chromosome and the ratio of transitions to transversions for a pair of sequences.

### Transcriptome

Whole blood was collected using the PAXgene RNA tube, which is widely used for transcriptome analysis. After storage at  $-80^{\circ}\text{C}$ , total RNA was purified with PAXgene Blood RNA Kit (Qiagen) using QiaSymphony (Qiagen). Total RNA was reverse-transcribed using an oligo-dT primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina) for library preparation for sequencing with HiSeq 2500 Sequencing System. For the quality assurance, we randomly selected 11 samples in one batch (usually 48 samples) and checked an RNA integrity number (RIN) (or an RIN equivalent) using BioAnalyzer or Tape Station (both from Agilent Technologies, Santa Clara, California, USA). The batch with RIN (or an RIN equivalent) higher than 7.0 for all tested samples was used for the downstream analysis. The minimum threshold for the total sequence reads for each sample was set to 30 million. For computing a series of QC metrics for RNA-seq data, RNA-SeQC was used to check the quality of sequence reads.<sup>34</sup>

### Plasma and urine metabolome

#### Nuclear magnetic resonance (NMR) spectroscopy

All NMR measurements for metabolome analysis were conducted at 298 K on a Bruker Avance 600 MHz spectrometer equipped with a SampleJet sample changer (Bruker, Billerica, Massachusetts, USA).<sup>35</sup> Standard 1-dimensional nuclear Overhauser enhancement spectroscopy and Carr-Purcell-Meiboom-Gill spectra were obtained for each plasma or urine sample. All spectra for plasma or urine samples were acquired using 16 scans and 32 k of complex data points. All data were analysed using the TopSpin 3.5 (Bruker) and Chenomx NMR Suite 8.2 (Chenomx, Edmonton, Alberta, Canada) programmes. All spectra were referenced to an internal standard (DSS-d6). As necessary, those spectra were aligned using hierarchical cluster-based peak alignment method, which is implemented as an R package called 'speaq'.<sup>36</sup>

#### Gas chromatography-tandem mass spectrometry (GC-MS/MS)

Sample preparation for plasma and urine (50  $\mu\text{L}$  each) was performed using a Microlab STARlet robot system (Hamilton, Reno, Nevada, USA) followed by the methods previously reported by Nishiumi *et al.*<sup>37 38</sup> The resulting deproteinised and derivatised supernatant (1  $\mu\text{L}$ ) was subjected to GC-MS/MS, performed on a GC-MS TQ-8040 system (Shimadzu, Kyoto, Japan). The compound separation was performed using a fused silica capillary column (BPX-5; 30 m $\times$ 0.25 mm inner diameter; film thickness: 0.25  $\mu\text{m}$ ; Shimadzu). Metabolite detection was performed using Smart Metabolites Database (Shimadzu) that contained the relevant multiple reaction monitoring (MRM) method file and data regarding the GC analytical conditions, MRM parameters and retention index employed for the metabolite measurement. The database used in this study included data on 475 peaks from 334 metabolites. All peaks of metabolites detected from each sample was annotated and analysed using Traverse

MS (Reifys, Tokyo, Japan). Then, two types of normalisation were performed to these annotated metabolites. The first normalisation was performed using the peak of 2-isopropylmalic acid as an internal standard, which was added to each sample before analysis with GC-MS/MS. Then the second normalisation was performed using QC samples, which were injected after every 12 study samples according to the reference quality control (RQC) normalisation methods.<sup>39</sup> Normalised values of each metabolite in the QC samples were assessed by calculating coefficients of variation (CVs), and metabolites with CVs over 20% were eliminated.

### Oral microbiome

Analysis of oral microbiome was conducted by previously reported protocols.<sup>40</sup> In brief, saliva was collected in a 50 mL tube. Dental plaque was sampled by participants by brushing teeth with a sterilised toothbrush, and then suspending it in 0.5 mL Tris-EDTA for collection. Both samples were stored at  $-80^{\circ}\text{C}$  until the time of processing. DNA was extracted from saliva and dental plaque by standard glass bead-based homogenisation and subsequent purification with a silica-membrane spin column using PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, California, USA). DNA was eluted from the spin column with 30  $\mu\text{L}$  RNase-free water (Takara Bio, Inc., Shiga, Japan), and stored at  $-20^{\circ}\text{C}$  after determining the amount and purity of DNA with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware, USA). Using DNA extracted from saliva or dental plaque as a template, a part of the V4 variable region of the bacterial 16S rRNA gene was amplified by two-step PCR. Tag-indexed PCR products thus obtained were subjected to multiplex amplicon sequencing using MiSeq System with MiSeq Sequencing Reagent Kit V.3 (Illumina) according to the manufacturer's instructions. For the quality assurance, the minimum threshold of the total sequence reads for each sample was set to ten thousands, and the principal component analysis was used to eliminate outliers.

### Outcomes

The following obstetric complications represented the primary outcomes. Gestational age was confirmed by measuring fetal crown rump length from 9 weeks to 13 weeks of gestation using transvaginal ultrasound. HDP was defined as gestational hypertension, pre-eclampsia, superimposed preeclampsia or chronic hypertension.<sup>41 42</sup> Preterm birth was defined as spontaneous preterm labour, medically induced preterm labour or preterm premature rupture of membranes resulting in preterm birth at less than 37 weeks of gestation. GDM was diagnosed according to the International Association of the Diabetes and Pregnancy Study Groups criteria.<sup>43</sup> The secondary outcomes were maternal body weight, blood pressure, physical activity, lifestyle changes, perinatal mental disorders, fetal growth, fetal movement and birth weight.

### Sample size calculation

At this time, there is little reliable evidence to demonstrate how time-dependent trends of longitudinal dense data would differ by pregnancy outcomes. Therefore, a priori sample size calculation is not provided in the present study. However, considering that one of the main purposes of the MLOG study is to explore the relationship between patterns of longitudinal home blood pressure and the onset of HDP, we estimated a required sample size as follows. Based on the HDP incidence of approximately 10% at Tohoku University Hospital, with a statistical power of 90% and a significance level of 5%, a sample of 250 participants is required to detect a 5 mm Hg difference in average home blood pressure (with a 7 mm Hg SD) in the HDP group. To allow for 15% attrition and withdrawals during pregnancy, a minimum of 300 participants at baseline was required.

### Statistical analysis of longitudinal lifelog data

One of the major advantages of the MLOG study is the dense information for each participant. Especially, time points for lifelog data collection are highly dense for each participant. For these datasets, per-person analysis of dynamic relationships between variables can be applied.<sup>44</sup> Vector autoregressive modelling is a promising solution to find the predicates for each outcome. In addition, the Granger causality test can elucidate the temporal ordering of dynamic relationship between two or more variables and indicate putative causal associations.<sup>45</sup> Some types of lifelog data were generated automatically; the others were manually input. We will first detect outlier data points, depending on the type of each lifelog, and eliminate them. The missing time-series lifelog data, ranging in 15%–33% of the total data points, would be imputed using the EM-imputation algorithm, for example, Amelia library,<sup>46</sup> after normalising the data by data transformation if required. For downstream analysis, the data might be collapsed with time scale, for example, taking trimmed mean or median for each week, month or trimester.

### Statistical analysis of multiomics data

The present study allows combination of longitudinal lifelog data with multiomics data. In contrast to single omics analysis, the multiomics analysis would reveal the complicated interactions between one and another. However, the sample size for multiomics analysis is usually relatively small. Dimension reduction via unsupervised or supervised learning for each omics data would be key ingredients to derive meaningful patterns from high dimensional data sets. Also, obtaining low dimensional representations provides a mean to deal with the multiple testing problem by decreasing number of statistical tests. For gene expression data, surrogate variable analysis<sup>47</sup> and sparse factor analysis<sup>48</sup> are frequently used to capture unknown batch effects in advance to expression quantitative trait locus (eQTL) analysis. The extracted factors can be removed from raw expression data to increase power for detecting associated genes.<sup>49</sup> Several unsupervised clustering methods<sup>50–52</sup> would be also applicable to obtain

hidden patterns from dense time-course lifelog measurements, which might be related to pregnancy complications. Recently developed multiview factor analysis approaches<sup>53 54</sup> have been used to integrate heterogeneous omics data to identify essential components to distinguish disease subtypes from few hundreds of samples. This line of approach would be a promising way to characterise biological status such as gestational age and to predict clinical outcomes such as spontaneous preterm birth.

Standard analyses would be also applicable for the selected variables and extracted factors (features). The association of outcomes with each feature will be analysed using statistical hypothesis tests such as Welch's t-test, Fisher's exact test, the  $\chi^2$  test and others as appropriate. Multiple logistic regression modelling will be used to adjust for confounders and to assess whether each feature or combination of features can be used to predict outcomes. Stepwise selection algorithms or regularised algorithms (eg, Least absolute shrinkage and selection operator (LASSO), ridge regression or elastic net) will be used to select the optimal number of contributing features that maximise the predictive power using the leave-1-out cross-validation or K-fold cross-validation methods.

Individual genetic features may have an effect on outcomes; therefore, some aggregated genetic risk score should be included in the prediction model. For example, SNVs, including rare variants in or around a chromosome region of a known or estimated risk gene, could be aggregated by considering their impacts on biological function of the gene or their minor allele frequencies in the population. However, this study is limited in the number of study participants, and the aggregated risk score might therefore contribute only slightly to the predictive power. To create a more reliable risk score, the estimates from other large-scale cohort data using polygenic score tools, for example, PRSice,<sup>55</sup> could be used for this study.

## FINDINGS TO DATE

### Clinical background

A total of 302 women were enrolled, and the mean gestational weeks of recruitment was 16.4±4.9 weeks (mean±SD). A total of 285 participants have been followed up to delivery; their baseline clinical characteristics are described in table 1. The mean maternal age at delivery was 33.3±4.9 years. As for educational levels, 62% of the participants were high school graduates with or without vocational college education, and 21% had a college degree. The majority were employed (65%) in early pregnancy, and about 40% had a high household income (over 6million yen per year). Approximately 42% of the participants were over 35 years of age, 51% were parous and 22% were overweight or obese by their prepregnancy BMIs ( $\geq 25$  kg/m<sup>2</sup>). Overall, 8.4% of the participants had HDP, and 5.6% underwent spontaneous preterm birth. On average, infants were delivered at 38.0±2.3 weeks of gestation with a mean birth weight of 2907±572 g. The rate of low birth weight was 18%. Mean gestational weeks of the first and second blood sampling

**Table 1** Participant characteristics

Characteristics	Value
<b>Maternal (n=285)</b>	
Age at delivery, years, mean (SD)	33.3 (±4.9)
<b>Age at delivery, years, n (%)</b>	
20–24	12 (4.2)
25–29	45 (15.8)
30–34	107 (37.5)
35–39	90 (31.6)
40–44	30 (10.5)
45–49	1 (0.4)
<b>Education (n=81), n (%)</b>	
Elementary school/junior high school	5 (6.2)
High school	35 (43.2)
Vocational college	23 (28.4)
College degree and above	17 (21.0)
Others	1 (1.2)
Data not available	204
<b>Occupation (n=270), n (%)</b>	
Housewife or unemployed	93 (34.4)
Employed	175 (64.8)
Student	2 (0.7)
<b>Annual household income, yen (n=248), n (%)</b>	
<2million	17 (6.9)
2–4million	59 (23.8)
4–6million	73 (29.4)
6–8million	51 (20.6)
8–10million	22 (8.9)
>10million	26 (10.5)
<b>Parity, n (%)</b>	
0	140 (49.1)
1	93 (32.6)
≥2	52 (18.2)
Prepregnancy BMI*, kg/m <sup>2</sup> , mean (SD)	22.7 (±5.1)
<b>Prepregnancy BMI, kg/m<sup>2</sup>, n (%)</b>	
<18.5	36 (12.6)
18.5–24.9	186 (65.3)
25.0–29.9	34 (11.9)
≥30.0	29 (10.2)
Gestational weeks at delivery, mean (SD)	38.0 (±2.3)
<b>Mode of delivery, n (%)</b>	
Non-caesarean	179 (62.8)
Caesarean	106 (37.2)
<b>Pregnancy complication, n (%)</b>	
Hypertensive disorder of pregnancy	24 (8.4)
Spontaneous preterm birth	16 (5.6)

Continued

Table 1 Continued

Characteristics	Value
Neonatal (n=300)	
Birth weight, g, mean (SD)	2907 ( $\pm$ 572)
Sex, n (%)	
Male	168 (56)
Female	132 (44)
Low birth weight (<2500 g), n (%)	54 (18)

\*BMI, body mass index.

were  $17.0\pm 5.0$  and  $27.5\pm 2.5$ , respectively. The third blood sampling was performed at  $31.1\pm 3.0$  days after delivery on average. The length of enrolment ranged from 90 days to 396 days with a mean of  $216\pm 61$  days.

### Data acquisition

The percentage of data uploads as of June 2017 was calculated for the 285 final study participants. For each lifelog item, the upload rate for each participant was calculated from the total number of days of actual uploads divided by the number of days from enrolment to delivery. The mean upload rate for each lifelog item was 85.3% (physical activity), 82.1% (body weight), 80.4% (body temperature), 78.0% (morning home blood pressure), 71.6% (evening home blood pressure), 83.5% (sleep quality), 82.1% (condition of stool, severity of pain, severity of nausea, uterine contractions and palpitations) and 67.4% (fetal movement) (figure 3).

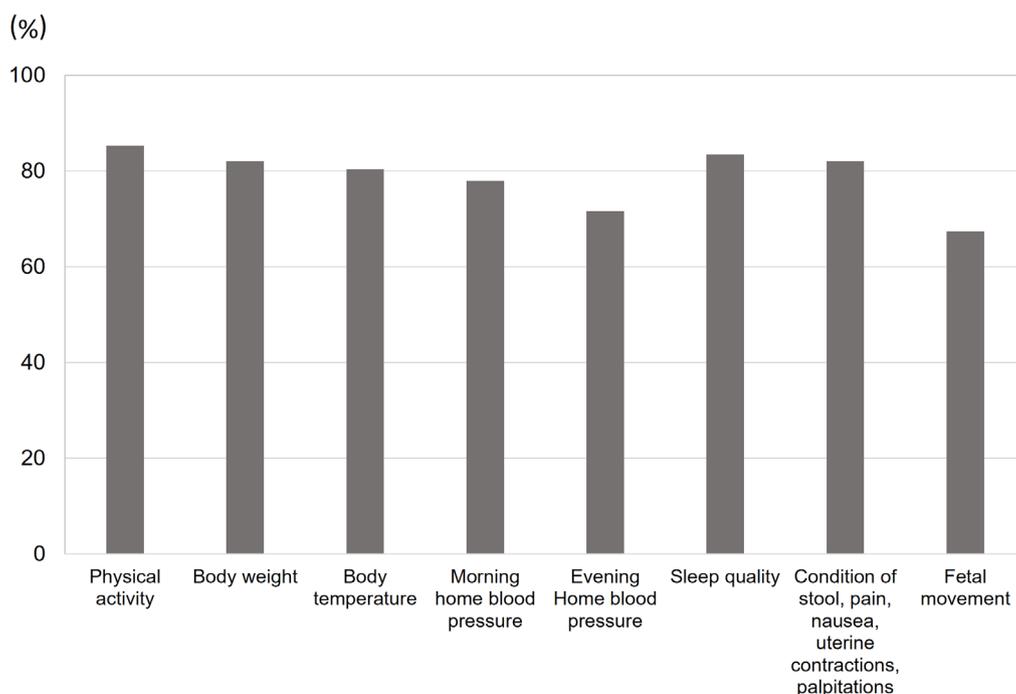
### Number of data points

The total number of collected data points as of June 2017 was calculated for the 285 final study participants. The approximate number of registered data points was 86 000 for body weight, 324 000 points for home diastolic and systolic blood pressure, 86 000 for physical activity and 74 000 for body temperature. When physical conditions such as stool condition, severity of pain and fetal movement were combined, the total number of data points was over 6million.

### STRENGTHS AND LIMITATIONS

Herein, we have described the rationale, design, objective, data collection methods and interim results of the MLOG study. The study was launched in September 2016, and baseline data collection ended in June 2017. A total of 285 participants uploaded lifelog data throughout pregnancy with a high data acquisition rate and over 6million total data points. Biospecimens for multiomics analysis were satisfactorily collected and all tracked by LIMS.

There are three noteworthy features in the MLOG study. First, it is a prospective add-on cohort study based on TMM BirThree Cohort Study, with a full series of epidemiological data and a highly structured follow-up system for mothers, newborns and families.<sup>8</sup> Second, we have successfully collected longitudinal, continuous, individual lifelog data with a high acquisition rate, which will enable us to assess dynamic changes in physiological conditions throughout pregnancy. Third, multiomics data will make it possible to fully understand the complex mechanisms of multifactorial pregnancy-related diseases and to overcome the unpredictability of these complications.



**Figure 3** Data acquisition rate. The mean data upload rate of specific measures was calculated from the total number of days of actual uploads divided by the number of days from enrolment to delivery for each participant.

Prediction models using clinical and epidemiological information and circulating factors for pregnancy-related diseases have been developed extensively,<sup>56</sup> and risk-assessment approaches using clinical information have also been developed.<sup>57,58</sup> However, there is a lack of evidence for the benefits of these predictive models for routine clinical use.<sup>59</sup> Once the likelihood of a pregnancy-related disorder is estimated with high sensitivity and specificity, evidence-based clinical interventions could reduce the rate of maternal and neonatal morbidity and mortality.<sup>60</sup> Therefore, an early-prediction algorithm that can be used with a high level of confidence is needed to obtain better outcomes for patients with pregnancy complications.

Recently, several studies of sample sizes comparable with ours exploiting lifelog or multiomics data were reported. One of the studies analysed lifelog and multiomics data, collected from 108 individuals at three time points during a 9-month period.<sup>61</sup> In their study, several remarkable relationships were identified among physiological and multiomics data through integrated analyses. Another study investigated genome-wide associations between genetic variants and gene expression levels across 44 human tissues from a few hundreds of postmortem donors.<sup>49</sup> They studied both *cis*-eQTL (within 1Mb of target-gene transcription start sites) and *trans*-eQTLs (more distant from target genes or on other chromosomes) with 350 whole blood samples and thereby identified 5862 *cis*-eQTL and one *trans*-eQTL associations. These previous studies indicate that our time-course high-resolution reference catalogue with 285 pregnant women would be well applicable to high-dimensional data analyses such as searches for quantitative trait loci and molecular risk markers.

Potential limitation of the present study is participant recruitment only at Tohoku University Hospital that is one of the tertiary hospitals in Miyagi Prefecture for high-risk populations. Therefore, the sample size is limited, and the results might not be applicable to the general populations. Inclusion criteria of the present study limited the eligibility to pregnant women with age >20 years and the ability to access the internet using a smartphone. Therefore, results of the present study might not be applicable to pregnancies with lower coverage of smartphone use.

Hopefully, our study will result in the development of a novel stratification model for pregnancy-related diseases employing multiomics and lifelog data.

The MLOG study will enable us to construct a time-course high-resolution reference catalogue of wellness and multiomics data from pregnant women and thereby develop a personalised predictive model for pregnancy complications. Progressive data sharing and collaborative studies would make it possible to establish a standardised early-prediction method through large clinical trials.

## COLLABORATION

We are very much interested in collaborating with other research groups and are open for specific and detailed proposals approved by the institutional ethical review

committee. We are planning to share the full data of the MLOG study in the TMM biobank<sup>8</sup> by the end of 2022, and a portion of the data have been distributed to researchers approved by the Sample and Data Access Committee of the biobank.

## Author affiliations

<sup>1</sup>Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan

<sup>2</sup>Tohoku University Graduate School of Medicine, Sendai, Japan

<sup>3</sup>Research Laboratories, NTT DoCoMo, Inc, Yokosuka, Japan

<sup>4</sup>Tohoku Medical and Pharmaceutical University, Sendai, Japan

<sup>5</sup>Radiation Effects Research Foundation, Hiroshima, Japan

<sup>6</sup>Tohoku University Hospital, Sendai, Japan

**Acknowledgements** The authors would like to thank all the MLOG study participants, the staff of the Tohoku Medical Megabank Organization, Tohoku University (a full list of members is available at <http://www.megabank.tohoku.ac.jp/english/a161201/>) and the Department of Obstetrics and Gynecology, Tohoku University Hospital, for their efforts and contributions. The MLOG study group also included Chika Igarashi, Motoko Ishida, Yumiko Ishii, Hiroko Yamamoto, Akiko Akama, Kaori Noro, Miyuki Ozawa, Yuka Narita, Junko Yusa, Miwa Meguro, Michiyo Sato, Miyuki Watanabe, Mai Tomizuka, Miika Hotta, Naomi Matsukawa, Makiko Sumii, Ayako Okumoto, Yukie Oguma, Ryoko Otokozaawa, Toshiya Hatanaka, Sho Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina and Daisuke Inoue.

**Contributors** JS, DO, RY, TY, HM, OT, SKuri, NY, SH and MN were involved in initial stages of the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH and MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM, OT and SKuri: recruitment and sample collection. DO, RY, TY, DS, TO, YT, YH, TFS, TM, JK, FK, TIT, SO, NM, SKo, OT and MN: sample analysis, data processing and statistical analysis. JS, HH, NF, NM, SKo, OT, SKuri, KK, SKure, NY, MY, SH and MN: advice and supervision of sample analysis. All authors have contributed to revision and have approved the final manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** The present study was supported by NTT DoCoMo, Inc, with a collaborative research agreement between NTT DoCoMo and ToMMo. This work was supported in part by the Tohoku Medical Megabank Project from the Japan Agency for Medical Research and Development and the Ministry of Education, Culture, Sports, Science and Technology.

**Competing interests** This study was funded by NTT DoCoMo, Inc. DO, TY and SH are employees of NTT DoCoMo, Inc.

**Patient consent** Obtained.

**Ethics approval** TMM BirThree Cohort Study was approved by the ethics committees of the Tohoku University (authorisation numbers, 2013-4-103 and 2017-4-010). The MLOG study was approved by the ethics committees of the Graduate School of Medicine (2014-1-704) and the Tohoku Medical Megabank Organization (2017-1-085), Tohoku University. Written informed consent was obtained from all participants.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** We are planning to share the full deidentified data of the MLOG study in the TMM biobank. Investigators interested in the MLOG study are encouraged to contact the corresponding authors, Dr Junichi Sugawara at [jsugawara@med.tohoku.ac.jp](mailto:jsugawara@med.tohoku.ac.jp) or Dr Masao Nagasaki at [nagasaki@megabank.tohoku.ac.jp](mailto:nagasaki@megabank.tohoku.ac.jp). Currently, no additional data are available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. *Diabetes Care* 2007;30(Suppl 2):S141-6.

2. Beck S, Wojdyla D, Say L, *et al.* The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bull World Health Organ* 2010;88:31–8.
3. Duley L. The global impact of pre-eclampsia and eclampsia. *Semin Perinatol* 2009;33:130–7.
4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980–2010: age-period-cohort analysis. *BMJ* 2013;347:f6564.
5. Waken RJ, de Las Fuentes L, Rao DC. A Review of the Genetics of Hypertension with a Focus on Gene-Environment Interactions. *Curr Hypertens Rep* 2017;19:23.
6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia / eclampsia. In: *Chesley's Hypertensive Disorders in pregnancy*. 2990. London: Elsevier:51–72.
7. Li X, Dunn J, Salins D, *et al.* Digital Health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol* 2017;15:e2001402.
8. Kuriyama S, Yaegashi N, Nagami F, *et al.* The Tohoku medical megabank project: design and mission. *J Epidemiol* 2016;26:493–511.
9. Japan Society of Obstetrics and Gynecology. *Guideline for obstetrical practice in Japan*. Tokyo, Japan: Japan Society of Obstetrics and Gynecology, 2017:1–4.
10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the thermoneutral zone: a longitudinal study of normal human pregnancy. *Acta Physiol* 2011;201:467–74.
11. Metoki H, Ohkubo T, Watanabe Y, *et al.* Seasonal trends of blood pressure during pregnancy in Japan: the babies and their parents' longitudinal observation in Suzuki Memorial Hospital in Intrauterine Period study. *J Hypertens* 2008;26:2406–13.
12. Haugen M, Brantsæter AL, Winkvist A, *et al.* Associations of pre-pregnancy body mass index and gestational weight gain with pregnancy outcome and postpartum weight retention: a prospective observational cohort study. *BMC Pregnancy Childbirth* 2014;14:201.
13. Sorensen TK, Williams MA, Lee IM, *et al.* Recreational physical activity during pregnancy and risk of preeclampsia. *Hypertension* 2003;41:1273–80.
14. Reutrakul S, Zaidi N, Wroblewski K, *et al.* Sleep disturbances and their relationship to glucose tolerance in pregnancy. *Diabetes Care* 2011;34:2454–7.
15. Cornish J, Tan E, Teare J, *et al.* A meta-analysis on the influence of inflammatory bowel disease on pregnancy. *Gut* 2007;56:830–7.
16. Huxley RR. Nausea and vomiting in early pregnancy: its role in placental development. *Obstet Gynecol* 2000;95:779–82.
17. Holm Tveit JV, Saastad E, Stray-Pedersen B, *et al.* Maternal characteristics and pregnancy outcomes in women presenting with decreased fetal movements in late pregnancy. *Acta Obstet Gynecol Scand* 2009;88:1345–51.
18. Facchinetti F, Allais G, D'Amico R, *et al.* The relationship between headache and preeclampsia: a case-control study. *Eur J Obstet Gynecol Reprod Biol* 2005;121:143–8.
19. Iams JD, Newman RB, Thom EA, *et al.* National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. Frequency of uterine contractions and the risk of spontaneous preterm delivery. *N Engl J Med* 2002;346:250–5.
20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular system. *Int J Cardiol* 2005;98:179–89.
21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol* 1997;32:920–4.
22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical practice and clinical research. *Tech Coloproctol* 2001;5:163–4.
23. Longstreth GF, Thompson WG, Chey WD, *et al.* Functional bowel disorders. *Gastroenterology* 2006;130:1480–91.
24. Koren G, Boskovic R, Hard M, *et al.* Motherisk-PUQE (pregnancy-unique quantification of emesis and nausea) scoring system for nausea and vomiting of pregnancy. *Am J Obstet Gynecol* 2002;186:S228–31.
25. Koren G, Piwko C, Ahn E, *et al.* Validation studies of the Pregnancy Unique-Quantification of Emesis (PUQE) scores. *J Obstet Gynaecol* 2005;25:241–4.
26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. *Br Med J* 1976;1:1305–7.
27. Winje BA, Saastad E, Gunnes N, *et al.* Analysis of 'count-to-ten' fetal movement charts: a prospective cohort study. *BJOG* 2011;118:1229–38.
28. United States. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191. *US Statut Large* 1996;110:1936–2103.
29. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification rules under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; other modifications to the HIPAA rules. *Fed Regist* 2013;78:5565–702.
30. Personal Information Protection Commission. Amended Act on the Protection of Personal Information (Tentative Translation). 2017 [https://www.ppc.go.jp/files/pdf/Act\\_on\\_the\\_Protection\\_of\\_Personal\\_Information.pdf](https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.pdf)
31. Takai-Igarashi T, Kinoshita K, Nagasaki M, *et al.* Security controls in an integrated Biobank to protect privacy in data sharing: rationale and study design. *BMC Med Inform Decis Mak* 2017;17:100.
32. Katsuoka F, Yokozawa J, Tsuda K, *et al.* An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. *Anal Biochem* 2014;466:27–9.
33. Yamaguchi-Kabata Y, Nariai N, Kawai Y, *et al.* iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. *Hum Genome Var* 2015;2:15050.
34. DeLuca DS, Levin JZ, Sivachenko A, *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 2012;28:1530–2.
35. Koshiha S, Motoike I, Kojima K, *et al.* The structural origin of metabolic quantitative diversity. *Sci Rep* 2016;6:31463.
36. Vu TN, Valkenburg D, Smets K, *et al.* An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics* 2011;12:405.
37. Nishiumi S, Kobayashi T, Ikeda A, *et al.* A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS One* 2012;7:e40459.
38. Nishiumi S, Kobayashi T, Kawana S, *et al.* Investigations in the possibility of early detection of colorectal cancer by gas chromatography/triple-quadrupole mass spectrometry. *Oncotarget* 2017;8:17115–26.
39. Saigusa D, Okamura Y, Motoike IN, *et al.* Establishment of protocols for global metabolomics by LC-MS for biomarker discovery. *PLoS One* 2016;11:e0160555.
40. Sato Y, Yamagishi J, Yamashita R, *et al.* Inter-Individual Differences in the Oral Bacteriome Are Greater than Intra-Day Fluctuations in Individuals. *PLoS One* 2015;10:e0131607.
41. Brown MA, Magee LA, Kenny LC, *et al.* Hypertensive disorders of pregnancy: isshp classification, diagnosis, and management recommendations for international practice. *Hypertension* 2018;72:24–43.
42. Watanabe K, Naruse K, Tanaka K, *et al.* Outline of definition and classification of "Pregnancy induced Hypertension (PIH)". *Hypertension Research in Pregnancy* 2013;1:3–4.
43. Metzger BE, Gabbe SG, Persson B, *et al.* International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* 2010;33:676–82.
44. Gep B, Jenkins GM, Reinsel GC. *Time series analysis: forecasting and control*. 5th edn. New Jersey: Wiley, 2015.
45. Brandt PT, Williams JT. *Multiple time series models*. Thousand Oaks, CA: Sage Publications, 2007.
46. Honaker J, King G, Blackwell M. Amelia II: a program for missing data. *Journal of Statistical Software* 2011;45.
47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007;3:1724–35.
48. Stegle O, Parts L, Piipari M, *et al.* Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 2012;7:500–7.
49. Battle A, Brown CD, Engelhardt BE, *et al.* Genetic effects on gene expression across human tissues. *Nature* 2017;550:204–13.
50. Polgreen PM, Yang M, Kuntz JL, *et al.* Using oral vancomycin prescriptions as a proxy measure for Clostridium difficile infections: a spatial and time series analysis. *Infect Control Hosp Epidemiol* 2011;32:723–6.
51. McDowell IC, Manandhar D, Vockley CM, *et al.* Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLoS Comput Biol* 2018;14:e1005896.
52. Hensman J, Rattray M, Lawrence ND. Fast nonparametric clustering of structured time-series. *IEEE Trans Pattern Anal Mach Intell* 2015;37:383–93.
53. Rohart F, Gautier B, Singh A, *et al.* mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 2017;13:e1005752.
54. Argelaquet R, Velten B, Arno D, *et al.* Multi-Omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol* 2018;14:e8124.
55. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics* 2015;31:1466–8.

56. Wax JR, Cartin A, Pinette MG. Biophysical and biochemical screening for the risk of preterm labor: an update. *Clin Lab Med* 2016;36:369–83.
57. Al-Rubaie Z, Askie LM, Ray JG, *et al.* The performance of risk prediction models for pre-eclampsia using routinely collected maternal characteristics and comparison with models that include specialised tests and with clinical guideline decision rules: a systematic review. *BJOG* 2016;123:1441–52.
58. Koullali B, Oudijk MA, Nijman TA, *et al.* Risk assessment and management to prevent preterm birth. *Semin Fetal Neonatal Med* 2016;21:80–8.
59. Henderson JT, Thompson JH, Burda BU, *et al.* Preeclampsia screening: evidence report and systematic review for the US Preventive Services Task Force. *JAMA* 2017;317:1668–83.
60. Broekhuijsen K, van Baaren GJ, van Pampus MG, *et al.* Immediate delivery versus expectant monitoring for hypertensive disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an open-label, randomised controlled trial. *Lancet* 2015;385:2492–501.
61. Price ND, Magis AT, Earls JC, *et al.* A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat Biotechnol* 2017;35:747–56.