BMJ Open

Maternity Log Study: protocol for a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated Pregnancy

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025939
Article Type:	Cohort profile
Date Submitted by the Author:	11-Aug-2018

Complete List of Authors:	Sugawara, Junichi; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology Ochi, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Yamashita, Riu; Tohoku Medical Megabank Organization, Tohoku University Yamauchi, Takafumi; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University Wagata, Maiko; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University Tsunemoto, Yoshiki; Research Laboratories, NTT DOCOMO, INC. Harada, Yuki; Tohoku Medical Megabank Organization, Tohoku University Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University Katsuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University Ogishima, Soichi; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical and Pharmaceutical University Hashizume, Hiroaki; Tohoku Medical Megabank Organization, Tohoku University Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Minegishi, Naoko; Tohoku Medical Megabank Organization, Tohoku University Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University Tanabe, Osamu; Tohoku Medical Megabank Organization, Tohoku University Kuriyama, Shinichi; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University

	Kure, Shigeo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yaegashi, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yamamoto, Masayuki; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Hiyama, Satoshi; Research Laboratories, NTT DOCOMO, INC. Nagasaki, Masao; Tohoku Medical Megabank Organization, Tohoku University
Keywords:	lifelog, multi-omics analysis, prediction, complicated pregnancy

Maternity Log Study: protocol for a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated pregnancy

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

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

3 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

4 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536.

5 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aobaku, Sendai, Miyagi 981-8558, Japan.

Email addresses:

*Junichi Sugawara: jsugawara@med.tohoku.ac.jp

Daisuke Ochi: ochi@megabank.tohoku.ac.jp

Riu Yamashita: ryamasi@megabank.tohoku.ac.jp

Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp

Daisuke Saigusa: saigusa@m.tohoku.ac.jp

Maiko Wagata: wagata@med.tohoku.ac.jp

Mami Ishikuro: m_ishikuro@med.tohoku.ac.jp

Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp

Yuki Harada: harada@megabank.tohoku.ac.jp

Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp

Junko Kawashima: kawashima@dent.tohoku.ac.jp

Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp

Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp

Soichi Ogishima: ogishima@megabank.tohoku.ac.jp

Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp

Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp

Nobuo Fuse: fusen@megabank.tohoku.ac.jp

Naoko Minegishi: nmine@med.tohoku.ac.jp

Seizo Koshiba: koshiba@megabank.tohoku.ac.jp

Osamu Tanabe: otanabe@megabank.tohoku.ac.jp

Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp

Kengo Kinoshita: kengo@ecei.tohoku.ac.jp

Shigeo Kure: kure@med.tohoku.ac.jp

Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp

Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp

Satoshi Hiyama: hiyamas@nttdocomo.com

*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

*Corresponding authors:

Junichi Sugawara: jsugawara@med.tohoku.ac.jp

and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, 9808573, Sendai, Japan. Phone: +81-22-273-6283

Word count: 3931 words

Abstract

Introduction: Multifactorial diseases, including various complications of pregnancy, are caused by a complex interaction of genetic and environmental factors such as lifestyle and living environment. The evaluation of continuous lifestyle monitoring using healthcare devices provides information on latent physiologic changes prior to the onset of disease. We expect that monitoring these factors directly is more accurate than using conventional methods such as questionnaires. A prospective cohort study for pregnant women, the Maternity Log study (MLOG), was designed to construct a time-course high-resolution reference catalog of bioinformatic data in pregnancy and explore the associations between genomic and environmental factors and the onset of pregnancy complications using continuous lifestyle monitoring combined with multi-omics data on the genome, transcriptome, proteome, metabolome, and microbiome.

Methods and analysis: Pregnant women were recruited at the timing of first routine antenatal visits. 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 physiologic data, such as body weight, blood pressure, heart rate, and body temperature, using multiple home healthcare devices. Biospecimens, including maternal plasma, serum, urine, saliva, dental plaque, and cord blood, were collected for multi-omics analysis. This study is expected to elucidate the causal relationship between complicated pregnancy and maternal lifestyle and physiologic changes. Lifelog and multi-omics data will be used to construct a time-course high-resolution reference catalog of pregnancy. The reference

catalog will allow us to discover relationships among multi-dimensional phenotypes and novel risk markers in pregnancy for the future personalized early prediction of pregnancy complications.

Ethics and dissemination: This study was approved by the Tohoku Medical Megabank Organization, Tohoku University (2014-1-704 and 2017-1-085). Written informed consent was obtained from all participants.

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.

Mutli-omics data will make it possible to understand the complex mechanisms of multifactorial pregnancy-related diseases.

A time-course high-resolution reference catalog of wellness and multi-omics data will be informative to develop a personalized predictive model for pregnancy complications.

Further study with larger sample size is needed to validate a reference catalog of normal pregnancy and a prediction model of pregnancy complications.

INTRODUCTION

The incidence of pregnancy-related disorders, including hypertensive disorders of pregnancy (HDP), gestational diabetes mellitus (GDM), and preterm delivery has been increasing worldwide [1-4]. These multifactorial conditions are caused by an interaction of genetic factors and environmental factors [5,6]. Recent reports suggest that continuous lifestyle monitoring using wearable biosensors provides important information on latent physiologic changes that are exhibited prior to the onset of disease [7]. 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 physiologic data using multiple home healthcare devices. In addition, variety of biospecimens are collected for multi-omics analysis.

To the best of our knowledge, this study will be the first to integrate multi-omics data with objective data on environmental factors, including daily lifelog data, in pregnant women. Integrated information from the study will be utilized to discover the relationship among multi-dimensional phenotypes and novel risk markers for the future personalized early prediction of pregnancy complications.

METHODS AND ANALYSIS

Study setting

The aim of the MLOG study is to construct a time-course high-resolution reference catalog of bioinformatic data in pregnancy and thereby develop methods by which early prediction of obstetric complications, through integrated analysis of daily lifelogs and multi-omics data, *i.e.*, maternal genomes,

transcriptomes, metabolomes, and oral microbiomes.

The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-generation cohort study established by the Tohoku Medical Megabank Organization (TMM BirThree Cohort Study) [8] 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 Tohoku Medical Megabank Organization (ToMMo) [8].

Written informed consent was obtained from all participants by the genome medical research coordinators (GMRCs). This study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of ToMMo, Tohoku University (2014-1-704, 2017-1-085). This study was conducted under a collaborative research agreement with ToMMo, Tohoku University and NTT DoCoMo, Inc. (Tokyo, Japan).

Patient and public involvement

Patients and 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 September 2016. A flowchart of the recruitment process is shown in Figure 1. Patients who already

agreed to participate in the TMM BirThree Cohort Study were recruited to provide an additional informed consent for the MLOG study. A total of 302 women were enrolled. The inclusion criteria were age ≥ 20 years and the ability to access the internet using a smartphone in the Japanese language. Participants were excluded after enrollment 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 wearable 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 to 35 weeks, and every week from 36 weeks to delivery [9]. 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 3 times from each participant; the first sample was collected between 12 and 24 weeks of gestation, the second between 24 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°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, NJ, USA) at -80°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 blood was collected from the umbilical vein in a PAXgene® tube for storage at -80°C, and in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) for separation of plasma to be stored at -80°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°C until the time of analysis.

Saliva and dental plaque sampling

Samples of saliva and dental plaque were collected 3 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, NY, USA) and stored at -80°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°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, *i.e.*, body temperature [10], home blood pressure [11], body weight [12], physical activity (calorie expenditure) [13], as well as self-administered

information such as sleep quality [14], condition of stool [15], severity of nausea [16], fetal movement [17], severity of pain [18], uterine contractions [19], and palpitations [20]. 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 [21-23]; severity of nausea using the Pregnancy-Unique Quantification of Emesis and nausea (PUQE) score [24,25]; 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 [26,27].

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 diarrhea [21, 22], and its use has been spread widely to evaluate functional bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types according to cohesion and surface cracking [21, 22].

The PUQE score [24, 25] 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 h). The total score ranges from 3 (no symptoms) to 15, and higher scores are

correlated with increasing severity of nausea and vomiting [24, 25].

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 [26]. 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 9 movements [27]. 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 favorable 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 Co., Ltd., 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 meter (OMRON Healthcare Co., Ltd.) 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 Co., Ltd.) just after awakening. Physical activity was assessed using an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count steps and calculate calorie expenditure.

Clinical and epidemiological information

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

Database

A customized laboratory information management system (LIMS) was established to track all biospecimens. All data were transferred to the TMM integrated database after 2-step anonymization in a linkable fashion.
Data handling was strictly regulated under HIPAA (Health Insurance Portability and Accountability Act of 1996, United States Security and Privacy Rules) [28, 29] and the Act on the Protection of Personal Information [30]. Security control at our facility has been described previously [31].

Omics analysis

Whole-genome sequencing

To minimize amplification bias, we adopted a PCR-free library preparation method. After performing library quality control using the quantitative MiSeq method [32], libraries were sequenced on HiSeq 2500 Sequencing System (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We generated the sequencing data at over 12.5x coverage on average, and we identified variants using the alignment tool BWA-MEM (ver. 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 (ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's Variant Quality Score Recalibration (VQSR) 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.megabank.tohoku.ac.jp).

Transcriptome

Whole blood were collected using the PAXgene® RNA tube, which is widely used for transcriptome analysis. After storage at -80°C, total RNA was purified with PAXgene Blood RNA Kit® (Qiagen, Venlo, The Netherlands) using QiaSymphony® (Qiagen). The amount and quality of the total RNA was assessed with Bio Analyzer® or Tape Station® (both from Agilent Technologies, Santa Clara, CA, USA), and we only used RNA samples with an RNA integrity number (RIN) (or an RIN equivalent) higher than 7.0. Total RNA was reverse-transcribed using an oligo-dT primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.) for library preparation for

sequencing with HiSeq 2500 Sequencing System.

Plasma and urine metabolome

Nuclear magnetic resonance (NMR) spectroscopy

All NMR measurements for metabolome analysis were conducted at 298 K on a Brucker Avance 600 MHz spectrometer equipped with a SampleJet sample changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-Meiboom-Gill (CPMG) 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 analyzed using the TopSpin 3.5 (Bruker Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta, Canada) programs.

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

Sample preparation for plasma and urine (50 µL each) was performed using a Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the methods previously reported by Nishiumi [36, 37]. The resulting deproteinized and derivatized supernatant (1 µL) was subjected to GC-MS/MS, performed on a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound separation was performed using a fused silica capillary column (BPX-5; 30 m × 0.25 mm inner diameter; film thickness, 0.25 µm; Shimadzu Corp, Kyoto, Japan). Metabolite detection was performed using Smart Metabolites Database (Shimadzu Corp.) 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 analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan).

Oral Microbiome

Analysis of oral microbiome was conducted by previously reported protocols [36]. In brief, saliva was collected in a 25-mL tube. Dental plaque was sampled by participants themselves by brushing teeth with a sterilized toothbrush, and then suspended in saline for collection. Both samples were stored at -80°C until the time of processing. DNA was extracted from saliva and dental plaque by standard glass bead-based homogenization and subsequent purification with a silica-membrane spin-column using PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin column with 30-μL RNase-free water (Takara Bio, Inc., Shiga, Japan), and stored at -20°C after determining the amount and quality of DNA with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, 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 2-step PCR. Tag-indexed PCR products thus obtained were subjected to multiplex amplicon sequencing using MiSeq System and MiSeq Sequencing Reagent Kit, v3 (Illumina, Inc.) according to the manufacturer's instructions.

Outcomes

The following obstetric complications represented the primary outcomes. HDP

was classified as gestational hypertension, preeclampsia, superimposed preeclampsia, and chronic hypertension [37]. Spontaneous preterm birth was defined as spontaneous preterm labor 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 (IADPSG) criteria [38]. The secondary outcomes were maternal body weight, blood pressure, physical activity, lifestyle changes, perinatal mental disorders, fetal growth, fetal movement, and birth weight.

Statistical analysis

The association of outcomes with each factor will be analyzed using a statistical hypothesis test such as Welch's t-test, Fisher's exact test, the Chi-square test, and others as appropriate. Multiple logistic regression modelling will be used to adjust for confounders and to assess whether each factor or combination of factors can be used to predict outcomes. Stepwise selection algorithms or regularized algorithms (*e.g.,* LASSO, ridge regression, or elastic net) will be used to select the optimal number of contributing factors that maximize the predictive power using the leave-1-out cross validation or K-fold cross validation methods.

Individual genetic factors 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, we are 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, *e.g.*, PRSice [39], could be used for this study.

INTERIM RESULTS

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 delivered; their baseline clinical characteristics are described in Table 1. The mean maternal age at delivery was 33.3 ± 4.9 years. Approximately 42% of the participants were over 35 years of age, 51% were parous, and 22% were overweight or obese by their prepregnancy body mass indices (BMI ≥ 25 kg/m^2). Overall, 8.4% of 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 were 17.0 ± 5.0 and 27.5 ± 2.5, respectively. The third blood sampling was performed at 31.1 ± 3.0 days after delivery on average. The length of enrollment ranged from 90 to 396 days with a mean of 216 ± 61 days.

Table 1. Participant characteristics

Characteristic	Value
Maternal (n = 285)	
・Age at delivery, y, mean (SD)	33.3 (± 4.9)
・Age at delivery, y, n (%)	
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)
・Parity, n (%)	
0	140 (49.1)
1	93 (32.6)
≥ 2	52 (18.2)
・Prepregnancy BMI*, kg/m^2, mean (SD)	22.7 (± 5.1)
・Prepregnancy BMI, kg/m^2, n (%)	
< 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)
・Mode of delivery, n (%)	
Noncesarean	179 (62.8)
Cesarean	106 (37.2)
・Pregnancy complication, n (%)	
Hypertensive disorder of pregnancy	24 (8.4)
Spontaneous preterm birth	16 (5.6)
Neonatal (n = 300)	
・Birth weight, mean (SD)	2907 (± 572)
・Sex, n (%)	
Male	168 (56)
Female	132 (44)
・Low-birth weight (< 2500 g), n (%)	54 18)

*BMI, body mass index

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 days of actual upload divided by the number of days from enrollment to delivery. The mean upload rate for each lifelog item was 85.3% (steps and calorie), 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, 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 6 million.

DISCUSSION

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 6 million total data points. Biospecimens for multi-omics 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 the ToMMo BirThree cohort study, with a full series of epidemiological data and a highly structured follow-up system for mothers, newborns, and families [8]. Second, we have successfully collected longitudinal, continuous, individual lifelog data with a high acquisition rate, which will enable us to assess dynamic changes in physiologic conditions throughout pregnancy. Third, mutli-omics data will make it possible to fully understand the complex mechanisms of multifactorial pregnancy-related

diseases and to overcome the unpredictability of these complications.

Prediction models using clinical and epidemiological information and circulating factors for pregnancy-related diseases have been extensively developed [40], and risk-assessment approaches using clinical information have also been developed [41, 42]. However, there is a lack of evidence for the benefits of these predictive models for routine clinical use [43].

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 [44]. 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 multi-omics data were reported. One of the studies analyzed lifelog and multi-omics data, collected from 108 individuals at three time points during a nine-month period [45]. In their study, several remarkable relationships were identified among physiological and multi-omics 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 [46]. They studied both cis-eQTL (within 1 Mb 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 5,862 cis-eQTL and one trans-eQTL associations. These previous studies indicate that our time-course high-resolution reference catalog with 285 pregnant women would be well applicable to high-dimensional data analyses such as searches for quantitative trait loci and molecular risk markers.

Hopefully, our study will result in the development of a novel stratification model for pregnancy-related diseases employing multi-omics and lifelog data. The MLOG study will enable us to construct a time-course high-resolution reference catalog of wellness and multi-omics data from pregnant women and thereby develop a personalized predictive model for pregnancy complications. Progressive data sharing and collaborative studies would make it possible to establish a standardized early-prediction method through large clinical trials.

Author affiliations

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

3 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

4 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536.

5 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aobaku, Sendai, Miyagi 981-8558, 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, Mika Hotta, Naomi Matsukawa, Makiko Sumii, Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.

Contributors

JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH, MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM, OY, SKu: recruitment and sample collection. DO, RY, TY, DS, YT, YH, TFS, JK, FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data processing, and statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK, NY, MY, SH, MN: advice and supervision of sample analysis. All authors have contributed to revision and have approved the final manuscript.

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 (AMED) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Competing interests

This study was funded by NTT DoCoMo, Inc.

Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT DoCoMo, Inc. All other authors declare that they have no competing interests.

Ethics approval and consent to participate

The TMM BirThree cohort study was approved by the ethics committee of the Tohoku University (authorization numbers, 2013-4-103 and 2017-4-010). The MLOG study was approved by the Tohoku Medical Megabank Organization, Tohoku University (2014-1-704 and 2017-1-085). Written informed consent was obtained from all participants.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

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, Betran AP, Merialdi M, Requejo JH, 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: f 6564.

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. London: Elsevier; 2990: 51-72.

7. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schu ¨ssler-Fiorenza Rose SM, 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, Arai T, Kawaguchi Y, Osumi N, 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, Japan Society of Obstetrics and Gynecology, Tokyo, Japan, pp. 1–4, 2017 [in Japanese].

10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the thermoneutral zone: a longitudinal study of normal human pregnancy. Acta Physiol (Oxf). 2011; 201: 467-74.

11. Metoki H, Ohkubo T, Watanabe Y, Nishimura M, Sato Y, Kawaguchi M, Hara A, Hirose T, Obara T, Asayama K, Kikuya M, Yagihashi K, Matsubara Y, Okamura K, Mori S, Suzuki M, Imai Y; BOSHI Study Group. 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, Lissner L, Alexander J, Oftedal B, Magnus P, Meltzer HM. 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 Jun 11; 14: 201.

13. Sorensen TK, Williams MA, Lee IM, Dashow EE, Thompson ML, Luthy DA. Recreational physical activity during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-80.

14. Reutrakul S, Zaidi N, Wroblewski K, Kay HH, Ismail M, Ehrmann DA, Van Cauter E. Sleep disturbances and their relationship to glucose tolerance in pregnancy. Diabetes Care. 2011; 34: 2454-7.

15. Cornish J, Tan E, Teare J, Teoh TG, Rai R, Clark SK, Tekkis PP. 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, Børdahl PE, Frøen JF. 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, Benedetto C, Volpe A. 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, Goldenberg RL, Mueller-Heubach E, Moawad A, Sibai BM, Caritis SN, Miodovnik M, Paul RH, Dombrowski MP, Thurnau G, McNellis D; 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, Houghton LA, Mearin F, Spiller RC. Functional Bowel Disorders. Gastroenterology 2006; 130: 1480–91.

24. Koren G, Boskovic R, Hard M, Maltepe C, Navioz Y, Einarson A. 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, Boskovic R, Maltepe C, Einarson A, 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, Tveit JV, Stray-Pedersen B, Flenady V, et al. Analysis of 'count-to-ten' fetal movement charts: a prospective cohort study. BJOG. 2011;118: 1229-38.

28. 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.

29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191. US Statut Large. 1996; 110:1936-2103.

30. Amended Act on the Protection of Personal Information. https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.pdf.

31. Takai-Igarashi T, Kinoshita K, Nagasaki M, Ogishima S, Nakamura N, Nagase S, 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, Ito S, Pan X, Nagasaki M, et al. An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. Anal Biochem. 2014; 466: 27-9.

33. Koshiba S, Motoike I, Kojima K, Hasegawa T, Shirota M, Saito T, et al. The structural origin of metabolic quantitative diversity. Sci Rep. 2016; 6: 31463.

34. Nishiumi S, Kobayashi T, Ikeda A, Yoshie T, Kibi M, Izumi Y, et al. A novel serum metabolomics-based diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459.

35. Nishiumi S, Kobayashi T, Kawana S, Unno Y, Sakai T, Okamoto K, et al. Investigations in the possibility of early detection of colorectal cancer by gas chromatography/triple-quadrupole mass spectrometry. Oncotarget. 2017; 8, 17115-17126.

36. Sato Y, Yamagishi J, Yamashita R, Shinozaki N, Ye B, Yamada T, et al. Inter-Individual Differences in the Oral Bacteriome Are Greater than Intra-Day Fluctuations in Individuals. PLoS One. 2015;10: e0131607.

37. Watanabe K, Naruse K, Tanaka K, Metoki H, Suzuki Y. Outline of definition and classification of pregnancy induced hypertension (PIH). Hypertens Res

Pregnancy 2013; 1: 3–4.

38. IADPSG Consensus Panel: International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682.

39. #PRSice PRSice: Polygenic Risk Score software. Euesden J, Lewis CM, O'Reilly PF, Bioinformatics, 2015; 31:1466-8.

40. 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.

41. Al-Rubaie Z, Askie LM, Ray JG, Hudson HM, Lord SJ. 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; 123:1441-1452.

42. Koullali B, Oudijk MA, Nijman TA, Mol BW, Pajkrt E. Risk assessment and management to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21: 80-8.

43. Henderson JT, Thompson JH, Burda BU, Cantor A. Preeclampsia Screening: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA. 2017; 317: 1668-1683.

729

730 44. Broekhuijsen K, van Baaren GJ, van Pampus MG, Ganzevoort W, Sikkema
731 JM, Woiski MD, et al; HYPITAT-II Study Group. Immediate delivery versus
732 expectant monitoring for hypertensive disorders of pregnancy between 34
733 and 37 weeks of gestation (HYPITAT-II): an open-label, randomised
734 controlled trial. Lancet. 2015; 385: 2492-2501.

735

736 45. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, McDonald
737 DT, Kusebauch U, Moss CL, Zhou Y, Qin S, Moritz RL, Brogaard K, Omenn
738 GS, Lovejoy JC, Hood L. A wellness study of 108 individuals using personal,
739 dense, dynamic data clouds. Nat Biotechnol. 2017; 35: 747-756.

740

741 46. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center
742 (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis
743 Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund;
744 NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source
745 Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core
746 Resource—VARI; Brain Bank Repository—University of Miami Brain
747 Endowment Bank; Leidos Biomedical—Project Management; ELSI Study;
748 Genome Browser Data Integration &Visualization—EBI; Genome Browser Data
749 Integration &Visualization—UCSC Genomics Institute, University of California
750 Santa Cruz; Lead analysts:; Laboratory, Data Analysis &Coordinating Center
751 (LDACC):; NIH program management:; Biospecimen collection:; Pathology:;
752 eQTL manuscript working group:, Battle A, Brown CD, Engelhardt BE,
753 Montgomery SB. Genetic effects on gene expression across human tissues.

Nature. 2017; 550: 204-213.

Figure titles and legends

Figure 1. Flowchart of Maternity Log Study (MLOG) participants

Figure 2. Overview of the MLOG study protocol

A: Participant timeline for the MLOG study.

B: Physiologic information collected using healthcare devices. Specific measures were uploaded each day from the time of enrollment (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 enrollment (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 and 20 weeks of gestation, respectively.

Figure 3. Data acquisition rate

The mean data upload rate of specific measures was calculated from the total days of actual uploads divided by the number of days from enrollment to delivery in each participant.

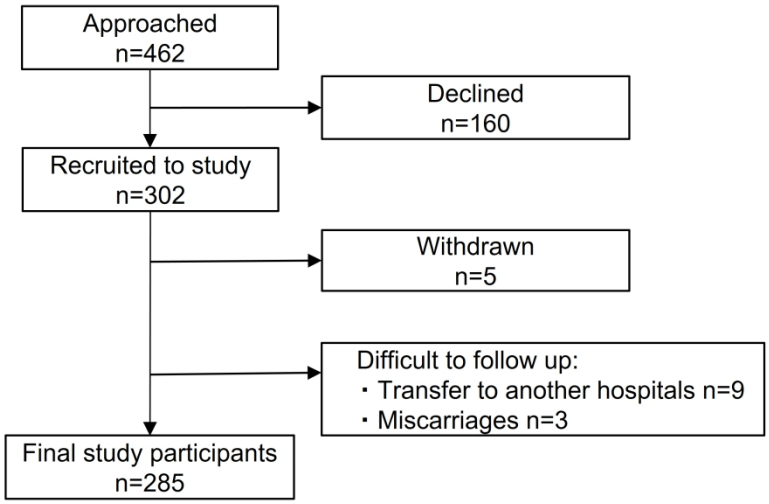

Figure 1.

254x190mm (300 x 300 DPI)

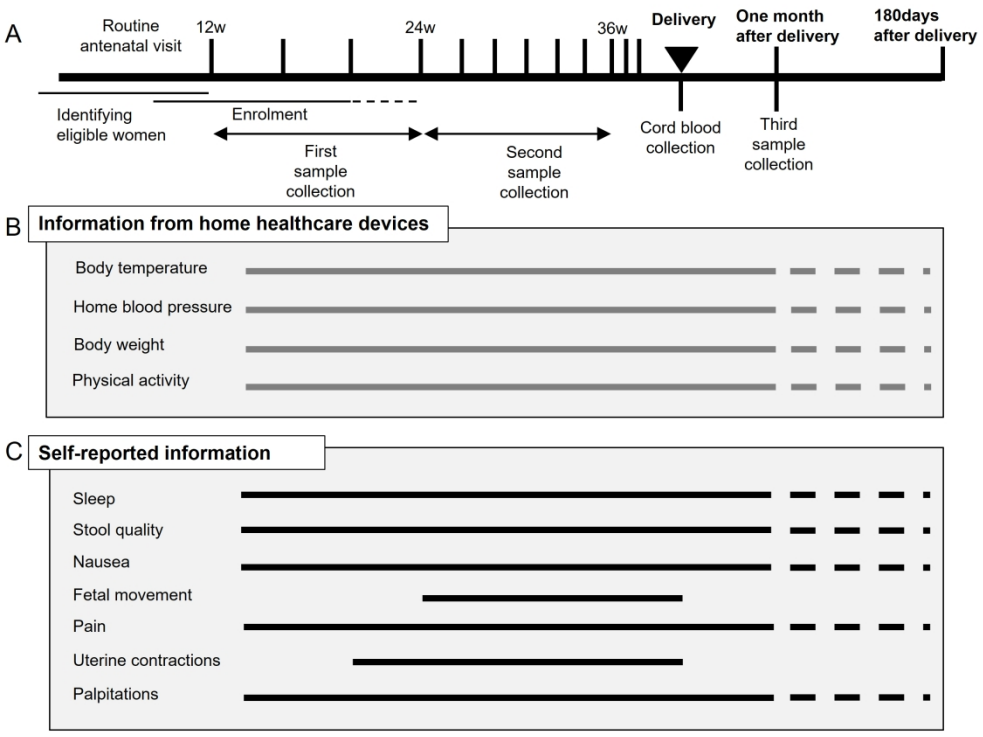

Figure 2.

254x190mm (300 x 300 DPI)

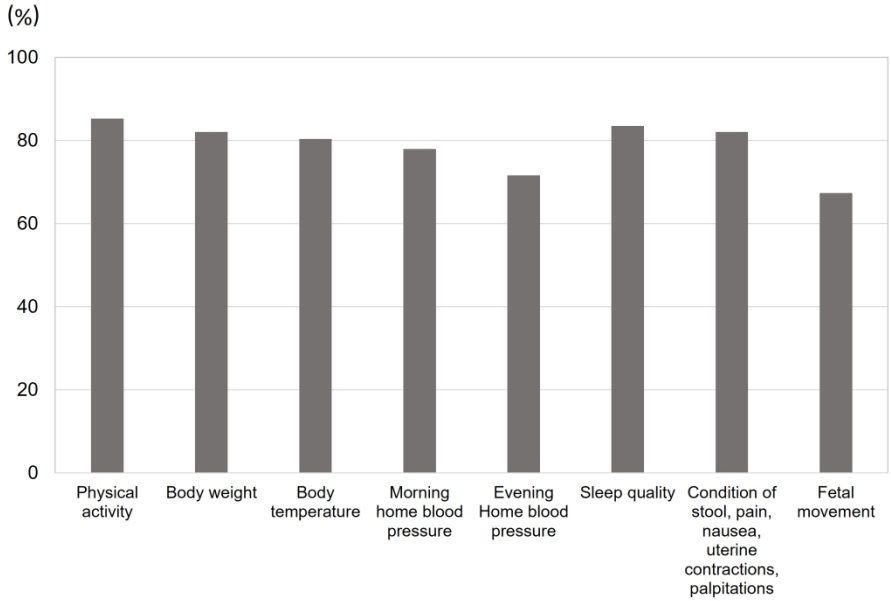

Figure 3.

254x190mm (300 x 300 DPI)

BMJ Open

Cohort Profile: Maternity Log Study: protocol for a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated Pregnancy

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025939.R1
Article Type:	Cohort profile
Date Submitted by the Author:	09-Nov-2018

Complete List of Authors:	Sugawara, Junichi; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology Ochi, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Yamashita, Riu; Tohoku Medical Megabank Organization, Tohoku University Yamauchi, Takafumi ; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University Wagata, Maiko; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Obara, Taku; Tohoku Medical Megabank Organization, Tohoku University Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University Tsunemoto, Yoshiki; Research Laboratories, NTT DOCOMO, INC. Harada, Yuki; Tohoku Medical Megabank Organization, Tohoku University Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University Mimori, Takahiro ; Tohoku Medical Megabank Organization, Tohoku University Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University Katsuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University Ogishima, Soichi; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical and Pharmaceutical University Hashizume, Hiroaki; Tohoku Medical Megabank Organization, Tohoku University Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Minegishi, Naoko; Tohoku Medical Megabank Organization, Tohoku University Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University

	Tanabe, Osamu; Tohoku Medical Megabank Organization, Tohoku University; Radiation Effects Research Foundation Kuriyama, Shinichi; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University Kure, Shigeo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yaegashi, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; TUH Yamamoto, Masayuki; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Hiyama, Satoshi; Research Laboratories, NTT DOCOMO, INC. Nagasaki, Masao; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine
Primary Subject Heading:	Obstetrics and gynaecology
Secondary Subject Heading:	Health informatics
Keywords:	lifelog, multi-omics analysis, prediction, complicated pregnancy

Cohort profile: Maternity Log Study: protocol for a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated pregnancy

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

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536.

4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai, Miyagi 981-8558, Japan.

5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan.

6 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-

8574, Japan.

Email addresses:

*Junichi Sugawara: jsugawara@med.tohoku.ac.jp

Daisuke Ochi: ochi@megabank.tohoku.ac.jp

Riu Yamashita: ryamasi@megabank.tohoku.ac.jp

Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp

Daisuke Saigusa: saigusa@m.tohoku.ac.jp

Maiko Wagata: wagata@med.tohoku.ac.jp

Taku Obara: obara-t@hosp.tohoku.ac.jp

Mami Ishikuro: m_ishikuro@med.tohoku.ac.jp

Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp

Yuki Harada: harada@megabank.tohoku.ac.jp

Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp

Takahiro Mimori: mimori@megabank.tohoku.ac.jp

Junko Kawashima: kawashima@dent.tohoku.ac.jp

Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp

Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp

Soichi Ogishima: ogishima@megabank.tohoku.ac.jp

Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp

Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp

Nobuo Fuse: fusen@megabank.tohoku.ac.jp

Naoko Minegishi: nmine@med.tohoku.ac.jp

Seizo Koshiba: koshiba@megabank.tohoku.ac.jp

Osamu Tanabe: otanabe@rerf.or.jp

Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp

Kengo Kinoshita: kengo@ecei.tohoku.ac.jp

Shigeo Kure: kure@med.tohoku.ac.jp

Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp

Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp

Satoshi Hiyama: hiyamas@nttdocomo.com

*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

*Corresponding authors:

Junichi Sugawara: jsugawara@med.tohoku.ac.jp

and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, 980-8573, Sendai, Japan. Phone: +81-22-273-6283

Word count: 5314 words

Key words: lifelog, multi-omics analysis, prediction, complicated pregnancy

Abstract

Purpose: A prospective cohort study for pregnant women, the Maternity Log study (MLOG), was designed to construct a time-course high-resolution reference catalog 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 labor, using continuous lifestyle monitoring combined with multi-omics 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 ≥ 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 physiologic 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 multi-omics analysis.

Future plans:

Lifelog and multi-omics data will be used to construct a time-course high-resolution reference catalog of pregnancy. The reference catalog will allow us to discover relationships among multi-dimensional phenotypes and novel risk markers in pregnancy for the future personalized 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. Multi-omics data will make it possible to understand the complex mechanisms of multifactorial pregnancy-related diseases.
- Potential limitations of the present study are as follows: 1) the limited sample size, and 2) participant recruitment only at a tertiary hospital for high-risk populations. Therefore, 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.
- Further study with a larger sample size of general populations is needed to validate a reference catalog of normal pregnancy and a prediction model of

pregnancy complications.

INTRODUCTION

The incidence of pregnancy-related disorders, including hypertensive disorders of pregnancy (HDP), gestational diabetes mellitus (GDM), and preterm delivery has been increasing worldwide [1-4]. These multifactorial conditions are caused by an interaction of genetic factors and environmental factors [5,6]. Recent reports suggest that continuous lifestyle monitoring using wearable biosensors provides important information on latent physiologic changes that are exhibited prior to the onset of disease [7]. 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 physiologic data using multiple home healthcare devices. In addition, a variety of biospecimens are collected for multi-omics analysis.

To the best of our knowledge, this study will be the first to integrate multi-omics 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 multi-dimensional phenotypes, and lead to identification of novel risk markers in pregnancy for the future personalized early prediction of pregnancy complications, e.g. hypertensive disorders of pregnancy, gestational diabetes, and preterm labor.

COHORT DESCRIPTION

Study setting

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

The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-generation cohort study established by the Tohoku Medical Megabank Organization (TMM BirThree Cohort Study) [8] 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 Tohoku Medical Megabank Organization (ToMMo) [8]. 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, a total of over 70,000 participants [8]. Written informed consent was obtained from all participants by the genome medical research coordinators (GMRCs). The MLOG study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committees of Graduate School of Medicine (2014-1-704) and ToMMo (22017-1-085), Tohoku University under a collaborative research agreement among ToMMo, Tohoku University

and NTT DoCoMo, Inc. (Tokyo, Japan).

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 flowchart 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 ≥ 20 years and the ability to access the internet using a smartphone in the Japanese language. Participants were excluded after enrollment 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 to 35 weeks, and every week from 36 weeks to delivery [9]. 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 3 times from each participant; the first sample was collected between 12 and 24 weeks of gestation, the second between 24 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°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, NJ, USA) at -80°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°C, and in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) for separation of plasma to be stored at -80°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°C until the time of analysis.

Saliva and dental plaque sampling

Samples of saliva and dental plaque were collected 3 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, NY, USA) and stored at -80°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°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, *i.e.*, body temperature [10], home blood pressure [11], body weight [12], and physical activity (calorie expenditure) [13], as well as self-administered information such as sleep quality [14], condition of stool [15], severity of nausea [16], fetal movement [17], severity of pain [18], uterine contractions [19], and palpitations [20]. 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 [21-23]; severity of nausea using the Pregnancy-Unique Quantification of Emesis and nausea (PUQE) score [24,25]; 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 [26,27].

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 diarrhea [21, 22], and its use has been spread widely to evaluate functional bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types according to cohesion and surface cracking [21, 22].

The PUQE score [24, 25] 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 h). The total score ranges from 3 (no symptoms) to 15, and higher scores are correlated with increasing severity of nausea and vomiting [24, 25].

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 [26]. 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 9 movements [27]. 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 favorable 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 Co., Ltd., 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 meter (OMRON Healthcare Co., Ltd.) 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 Co., Ltd.) just after awakening. Physical activity was assessed using an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count steps and calculate calorie expenditure.

Clinical and epidemiological information

Baseline clinical information and maternal and neonatal outcomes (*e.g.*, 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 [8].

Database

A customized laboratory information management system (LIMS) was established to track all biospecimens. All data were transferred to the TMM integrated database after 2-step anonymisation in a linkable fashion.

Data handling was strictly regulated under HIPAA (Health Insurance Portability and Accountability Act of 1996, United States Security and Privacy Rules) [28, 29] and the Act on the Protection of Personal Information [30]. Security control at our facility has been described previously [31].

Omics analysis

Whole-genome sequencing

To minimize amplification bias, we adopted a PCR-free library preparation method. After performing library quality control using the quantitative MiSeq method [32], libraries were sequenced on HiSeq 2500 Sequencing System (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We generated the sequencing data at over 12.5x coverage on average, and we identified variants using the alignment tool BWA-MEM (ver. 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 (ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's Variant Quality Score Recalibration (VQSR) 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.megabank.tohoku.ac.jp) [33]. 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°C, total RNA was purified with PAXgene Blood RNA Kit® (Qiagen, Venlo, The Netherlands) using QiaSymphony® (Qiagen). Total RNA was reverse-transcribed using an oligo-dT primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.) 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, CA, 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 thirty millions. For computing a series of quality control metrics for RNA-seq data, RNA-SeQC was used to check the quality of sequence reads [34].

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 Corp., Billerica, MA, USA) [35]. Standard 1-dimensional nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-Meiboom-Gill (CPMG) 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 analyzed using the TopSpin 3.5 (Bruker Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta, Canada) programs. 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" [36].

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

Sample preparation for plasma and urine (50 µL each) was performed using a Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the methods previously reported by Nishiumi [37, 38]. The resulting deproteinized and derivatized supernatant (1 µL) was subjected to GC-MS/MS, performed on a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound separation was performed using a fused silica capillary column (BPX-5; 30 m × 0.25 mm inner diameter; film thickness, 0.25 µm; Shimadzu Corp, Kyoto, Japan). Metabolite detection was performed using Smart Metabolites Database (Shimadzu Corp.) 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 analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan). Then, two types of normalization were performed to these annotated metabolites. The first normalization 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 normalization was performed using quality control (QC) samples which were injected after every 12 study samples according to the RQC normalization methods [39]. Normalized 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 [40]. In brief, saliva was collected in a 50-mL tube. Dental plaque was sampled by participants by brushing teeth with a sterilized toothbrush, and then suspending it in 0.5 mL Tris-EDTA for collection. Both samples were stored at -80°C until the time of processing. DNA was extracted from saliva and dental plaque by standard glass bead-based homogenization and subsequent purification with a silica-membrane spin-column using PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin column with 30-μL RNase-free water (Takara Bio, Inc., Shiga, Japan), and stored at -20°C after determining the amount and purity of DNA with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, 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 2-step PCR. Tag-indexed PCR products thus obtained were subjected to multiplex amplicon sequencing using MiSeq System with MiSeq Sequencing Reagent Kit, v3 (Illumina, Inc.) 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 to 13 weeks of gestation using transvaginal ultrasound. HDP was defined as gestational hypertension, preeclampsia, superimposed preeclampsia, or chronic hypertension [41,42]. Preterm birth was defined as spontaneous preterm labor, medically induced preterm labor, 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 (IADPSG) criteria [43]. 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-mmHg difference in average home blood pressure (with a 7-mmHg standard deviation) 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 [44]. Vector autoregressive (VAR) modeling 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 [45]. 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 - e.g. Amelia library [46], after normalising the data by data transformation if required. For downstream

analysis, the data might be collapsed with time scale, e.g. taking trimmed mean or median for each week, month, or trimester.

Statistical analysis of multi-omics data

The present study allows combination of longitudinal lifelog data with multi-omics data. In contrast to single omics analysis, the multi-omics analysis would reveal the complicated interactions between one and another. However, the sample size for multi-omics 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 [47] and sparse factor analysis [48] 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 [49]. Several unsupervised clustering methods [50,51,52] would be also applicable to obtain hidden patterns from dense time-course lifelog measurements, which might be related to pregnancy complications. Recently developed multi-view factor analysis approaches [53,54] 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 characterize 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 analyzed using statistical hypothesis tests such as Welch's t-test, Fisher's exact test, the Chi-square test, and others as appropriate. Multiple logistic regression modeling 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 regularized algorithms (*e.g.,* LASSO, ridge regression, or elastic net) will be used to select the optimal number of contributing features that maximize 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, *e.g.*, PRSice [55], 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 6 million 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 body mass indices (BMI ≥ 25 kg/m^2). 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 were 17.0 ± 5.0 and 27.5 ± 2.5, respectively. The third blood sampling was performed at 31.1 ± 3.0 days after delivery on average. The length of enrollment ranged from 90 to 396 days with a mean of 216 ± 61 days.

Table 1. Participant characteristics

Characteristics	Value
Maternal (n = 285)	
· Age at delivery, y, mean (SD)	33.3 (± 4.9)
· Age at delivery, y, n (%)	
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)
· Education (n=81) n (%)	
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

- Occupation (n=270) n (%)

Housewife or unemployed	93 (34.4)
Employed	175 (64.8)
Student	2 (0.7)

- Annual household income, yen (n=248) n (%)

< 2 million	17 (6.9)
2-4 million	59 (23.8)
4-6 million	73 (29.4)
6-8 million	51 (20.6)
8-10 million	22 (8.9)
> 10 million	26 (10.5)

- Parity, n (%)

0	140 (49.1)
1	93 (32.6)
≥ 2	52 (18.2)

- Prepregnancy BMI*, kg/m^2, mean (SD) 22.7 (± 5.1)
- Prepregnancy BMI, kg/m^2, n (%)

< 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)
- Mode of delivery, n (%)

Noncesarean	179 (62.8)
Cesarean	106 (37.2)

- Pregnancy complication, n (%)

Hypertensive disorder of pregnancy	24 (8.4)
Spontaneous preterm birth	16 (5.6)

Neonatal (n = 300)	
· Birth weight, g, mean (SD)	2907 (± 572)
· Sex, n (%)	
Male	168 (56)
Female	132 (44)
· Low-birth weight (< 2500 g), n (%)	54 (18)

*BMI, body mass index

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 enrollment 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 6 million.

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 6 million total data points. Biospecimens for multi-omics 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 [8]. Second, we have successfully collected longitudinal, continuous, individual lifelog data with a high acquisition rate, which will enable us to assess dynamic changes in physiologic conditions throughout pregnancy. Third, multi-omics data will make it possible to fully understand the complex mechanisms of multifactorial pregnancy-related diseases and to overcome the unpredictability of these complications.

Prediction models using clinical and epidemiological information and circulating factors for pregnancy-related diseases have been developed extensively [56], and risk-assessment approaches using clinical information have also been developed [57, 58]. However, there is a lack of evidence for the benefits of these predictive models for routine clinical use [59]. 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 [60]. 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 multi-omics data were reported. One of the studies analyzed lifelog and multi-omics data, collected from 108 individuals at three time points during a nine-month period [61]. In their study, several remarkable relationships were identified among physiological and multi-omics 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 [49]. They studied both cis-eQTL (within 1 Mb 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 5,862 cis-eQTL and one trans-eQTL associations. These previous studies indicate that our time-course high-resolution reference catalog 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 multi-omics and lifelog data. The MLOG study will enable us to construct a time-course high-resolution

reference catalog of wellness and multi-omics data from pregnant women and thereby develop a personalized predictive model for pregnancy complications. Progressive data sharing and collaborative studies would make it possible to establish a standardized early-prediction method through large clinical trials.

Author affiliations

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536.

4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai, Miyagi 981-8558, Japan.

5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan.

6 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, 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, Mika Hotta, Naomi Matsukawa, Makiko Sumii, Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.

Contributors

JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH, MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM, OY, SKu: recruitment and sample collection. DO, RY, TY, DS, TO, YT, YH, TFS, TM, JK, FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data processing, and statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK, NY, MY, SH, 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 (AMED) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Competing interests

This study was funded by NTT DoCoMo, Inc.

Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT DoCoMo, Inc. All other authors declare that they have no competing interests.

Ethics approval and consent to participate

TMM BirThree Cohort Study was approved by the ethics committees of the Tohoku University (authorization 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

The datasets used during the current study are available from the corresponding authors on reasonable request.

REFERENCES

1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. Diabetes Care. 2007;30 Suppl 2: S141-6. doi: 10.2337/dc07-s206.

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. doi: 10.2471/BLT.08.062554.

3. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol. 2009; 33:130-7. doi: 10.1053/j.semperi.2009.02.010.

4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. BMJ. 2013;347: f 6564. doi: 10.1136/bmj.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. doi: 10.1007/s11906-017-0718-1.

6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia / eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:

Elsevier; 2990: 51-72.

7. Li X, Dunn J, Salins D, Zhou G, et al. Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health Related Information. PLoS Biol. 2017; 15: e2001402. doi: 10.1371/journal.pbio.2001402.

8. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank Project: Design and Mission. J Epidemiol. 2016; 26:493-511. doi: 10.2188/jea.JE20150268.

9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo, Japan, pp. 1–4, 2017 [in Japanese].

10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the thermoneutral zone: a longitudinal study of normal human pregnancy. Acta Physiol (Oxf). 2011; 201: 467-74. doi: 10.1111/j.1748-1716.2010.02228.x.

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. doi: 10.1097/HJH.0b013e32831364a7.

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 Jun 11; 14: 201. doi: 10.1186/1471-2393-14-201.

13. Sorensen TK, Williams MA, Lee IM, et al. Recreational physical activity during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-80. doi: 10.1161/01.HYP.0000072270.82815.91

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. doi: 10.2337/dc11-0780.

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. doi: 10.1136/gut.2006.108324.

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. doi: 10.3109/00016340903348375.

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. doi: 10.1016/j.ejogrb.2004.12.020.

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. doi: 10.1056/NEJMoa002868

20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular system. Int J Cardiol. 2005; 98: 179-89. doi: 10.1016/j.ijcard.2003.10.028.

21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. Scand J Gastroenterol. 1997; 32: 920-4.
doi: 10.3109/00365529709011203

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. doi: 10.1007/s101510100019

23. Longstreth GF, Thompson WG, Chey WD, et al. Functional Bowel Disorders. Gastroenterology 2006; 130: 1480–91. doi: 10.1053/j.gastro.2005.11.061

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. doi: 10.1080/01443610500060651

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. doi: 10.1111/j.1471-0528.2011.02993.x

28. 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.

29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191. US Statut Large. 1996; 110:1936-2103.

30. Amended Act on the Protection of Personal Information. 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. doi: 10.1186/s12911-017-0494-5

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. doi: 10.1016/j.ab.2014.08.015

33.Yamaguchi-Kabata Y, et al, Nariai N, Kawai Y, et al. iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. Hum Genome Var. 2015; 2:15050. doi: 10.1038/hgv.2015.50. https://www.ncbi.nlm.nih.gov/pubmed/27081555

34. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012; 28: 1530-1532. doi: 10.1093/bioinformatics/bts196

35. Koshiba S, Motoike I, Kojima K, et al. The structural origin of metabolic quantitative diversity. Sci Rep. 2016; 6: 31463. doi: 10.1038/srep31463

36. Vu TN, Valkenborg D, Smets K, et al. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. BMC Bioinformatics. 2011; 12: 405. doi: 10.1186/1471-2105-12-405.

37. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459. doi: 10.1371/journal.pone.0040459.

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, 17115-17126. doi: 10.18632/oncotarget.15081.

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(8): e0160555. doi: 10.1371/journal.pone.0160555.

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. doi: 10.1371/journal.pone.0131607.

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. doi: 10.1161/HYPERTENSIONAHA.117.10803.

42. Watanabe K, Naruse K, Tanaka K, et al. Outline of definition and classification of pregnancy induced hypertension (PIH). Hypertens Res Pregnancy 2013; 1: 3–4.

43. IADPSG Consensus Panel: International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification

of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682. doi: 10.2337/dc09-1848.

44. Box GEP, Jenkins GM, Reinsel GC. Time series Analysis: Forecasting and Control. 5th ed. 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, 45 (7) 2011.

47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007; 3: 1724-35. doi: 10.1371/journal.pgen.0030161

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. doi: 10.1038/nprot.2011.457.

49. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene expression across human tissues. Nature. 2017; 550: 204-213. doi: 10.1038/nature24277.

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. doi: 10.1086/660858.

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. doi: 10.1371/journal.pcbi.1005896.

52. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of Structured Time-Series. IEEE Trans Pattern Anal Mach Intell. 2015; 37: 383-93. doi: 10.1109/TPAMI.2014.2318711.

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. doi: 10.1371/journal.pcbi.1005752.

54. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018; 14: e8124. doi: 10.15252/msb.20178124.

55. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics, 2015; 31:1466-8. doi: 10.1093/bioinformatics/btu848.

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. doi:

912 10.1016/j.cll.2016.01.019.

913

914 57. Al-Rubaie Z, Askie LM, Ray JG, et al. The performance of risk prediction
915 models for pre-eclampsia using routinely collected maternal characteristics and
916 comparison with models that include specialised tests and with clinical guideline
917 decision rules: a systematic review. BJOG. 2016; 123:1441-1452. doi:
918 10.1111/1471-0528.

919

920 58. Koullali B, Oudijk MA, Nijman TA, et al. Risk assessment and management
921 to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21: 80-8. doi:
922 10.1016/j.siny.2016.01.005.

923

924 59. Henderson JT, Thompson JH, Burda BU, et al. Preeclampsia Screening:
925 Evidence Report and Systematic Review for the US Preventive Services Task
926 Force. JAMA. 2017; 317: 1668-1683. doi: 10.1001/jama.2016.18315.

927

928 60. Broekhuijsen K, van Baaren GJ, van Pampus MG, et al; HYPITAT-II Study
929 Group. Immediate delivery versus expectant monitoring for hypertensive
930 disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an
931 open-label, randomised controlled trial. Lancet. 2015; 385: 2492-2501. doi:
932 10.1016/S0140-6736(14)61998-X.

933

934 61. Price ND, Magis AT, Earls JC, et al. A wellness study of 108 individuals
935 using personal, dense, dynamic data clouds. Nat Biotechnol. 2017; 35: 747-756.
936 doi: 10.1038/nbt.3870.

FIGURE TITLES AND LEGENDS

Figure 1. Flowchart of Maternity Log Study (MLOG) participants

Figure 2. Overview of the MLOG study protocol

A: Participant timeline for the MLOG study.

B: Physiologic information collected using healthcare devices. Specific measures were uploaded each day from the time of enrollment (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 enrollment (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 and 20 weeks of gestation, respectively.

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 enrollment to delivery for each participant.

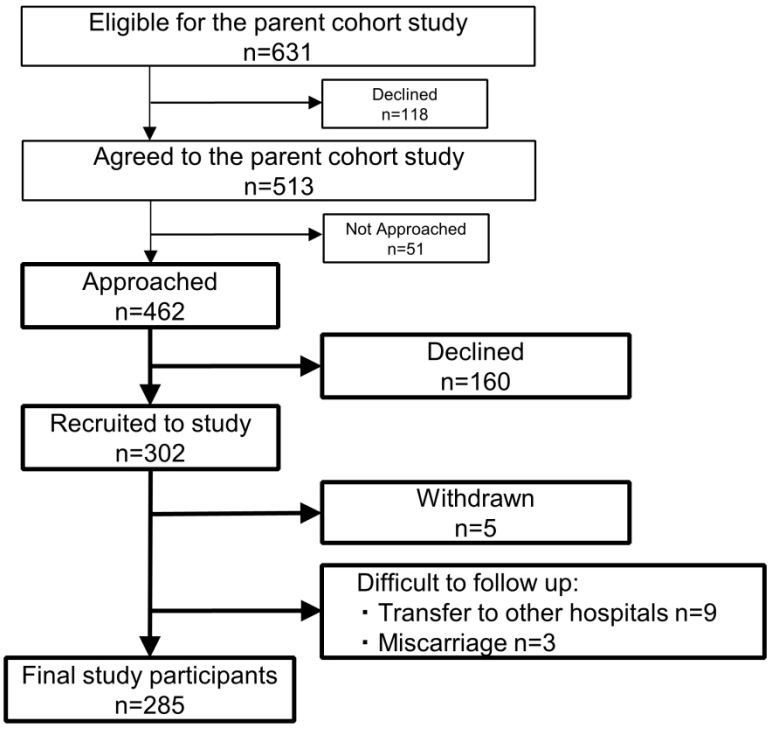

Figure 1.

254x190mm (300 x 300 DPI)

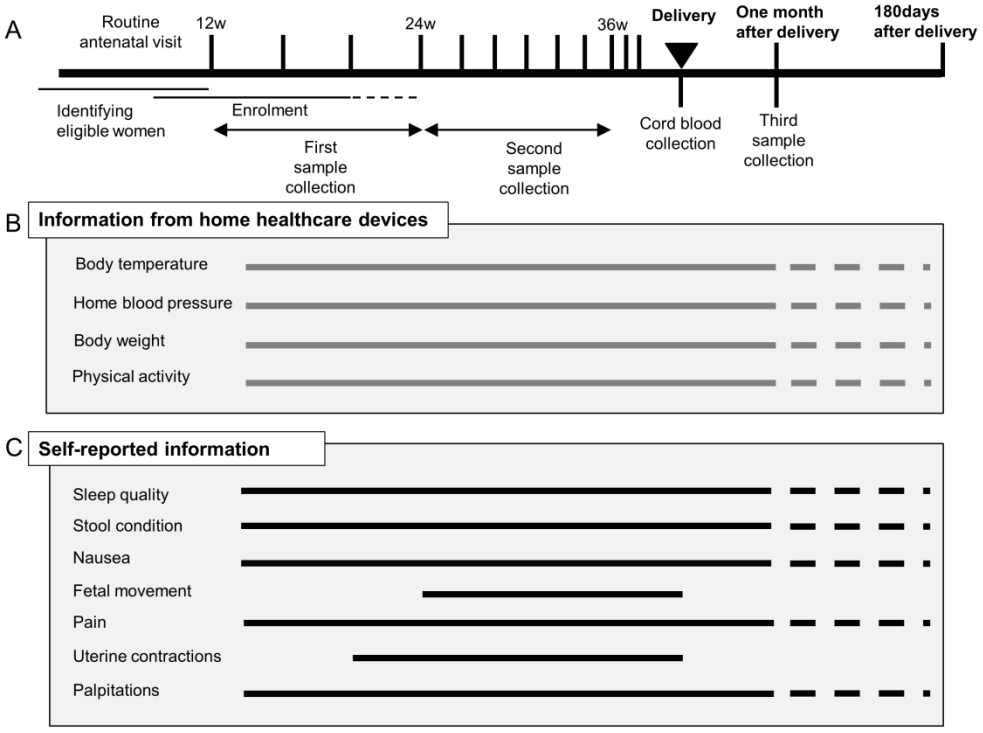

Figure 2.

254x190mm (300 x 300 DPI)

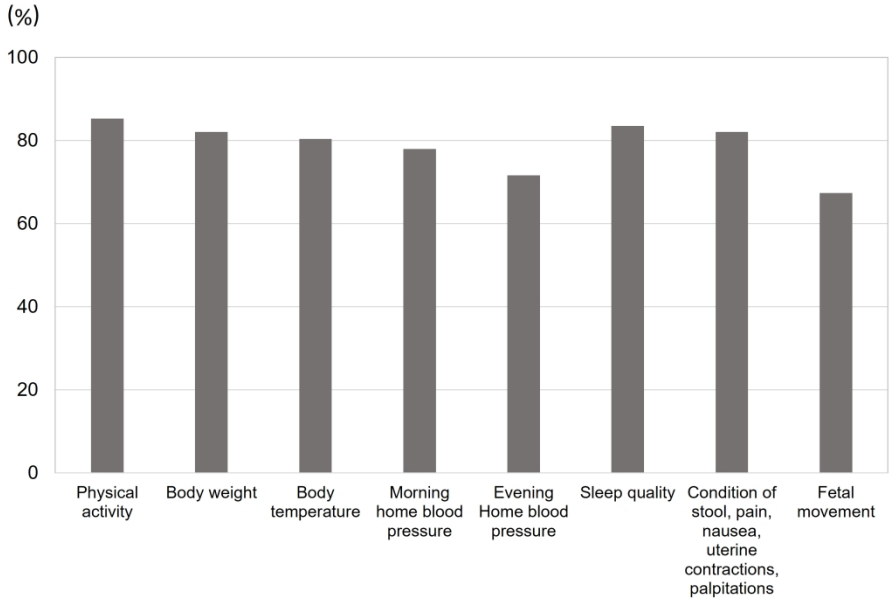

Figure 3.

254x190mm (300 x 300 DPI)

BMJ Open

Cohort Profile: Maternity Log Study: a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated Pregnancy

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-025939.R2
Article Type:	Cohort profile
Date Submitted by the Author:	20-Dec-2018

Complete List of Authors:	Sugawara, Junichi; Tohoku Medical Megabank Organization, Tohoku University, Feto-Maternal Medical Science; Tohoku University Graduate School of Medicine, Obstetrics and Gynecology Ochi, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Yamashita, Riu; Tohoku Medical Megabank Organization, Tohoku University Yamauchi, Takafumi ; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University Wagata, Maiko; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Obara, Taku; Tohoku Medical Megabank Organization, Tohoku University Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University Tsunemoto, Yoshiki; Research Laboratories, NTT DOCOMO, INC. Harada, Yuki; Tohoku Medical Megabank Organization, Tohoku University Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University Mimori, Takahiro ; Tohoku Medical Megabank Organization, Tohoku University Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University Katsuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University Ogishima, Soichi; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical and Pharmaceutical University Hashizume, Hiroaki; Tohoku Medical Megabank Organization, Tohoku University Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Minegishi, Naoko; Tohoku Medical Megabank Organization, Tohoku University Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University

	Tanabe, Osamu; Tohoku Medical Megabank Organization, Tohoku University; Radiation Effects Research Foundation Kuriyama, Shinichi; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University Kure, Shigeo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Yaegashi, Nobuo; Tohoku Medical Megabank Organization, Tohoku University; TUH Yamamoto, Masayuki; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Hiyama, Satoshi; Research Laboratories, NTT DOCOMO, INC. Nagasaki, Masao; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine
Primary Subject Heading:	Obstetrics and gynaecology
Secondary Subject Heading:	Health informatics
Keywords:	lifelog, multi-omics analysis, prediction, complicated pregnancy

Cohort profile: Maternity Log Study: a longitudinal lifelog monitoring and multi-omics analysis for the early prediction of complicated pregnancy

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

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536.

4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai, Miyagi 981-8558, Japan.

5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan.

6 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-

8574, Japan.

Email addresses:

*Junichi Sugawara: jsugawara@med.tohoku.ac.jp

Daisuke Ochi: ochi@megabank.tohoku.ac.jp

Riu Yamashita: ryamasi@megabank.tohoku.ac.jp

Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp

Daisuke Saigusa: saigusa@m.tohoku.ac.jp

Maiko Wagata: wagata@med.tohoku.ac.jp

Taku Obara: obara-t@hosp.tohoku.ac.jp

Mami Ishikuro: m_ishikuro@med.tohoku.ac.jp

Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp

Yuki Harada: harada@megabank.tohoku.ac.jp

Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp

Takahiro Mimori: mimori@megabank.tohoku.ac.jp

Junko Kawashima: kawashima@dent.tohoku.ac.jp

Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp

Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp

Soichi Ogishima: ogishima@megabank.tohoku.ac.jp

Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp

Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp

Nobuo Fuse: fusen@megabank.tohoku.ac.jp

Naoko Minegishi: nmine@med.tohoku.ac.jp

Seizo Koshiba: koshiba@megabank.tohoku.ac.jp

Osamu Tanabe: otanabe@rerf.or.jp

Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp

Kengo Kinoshita: kengo@ecei.tohoku.ac.jp

Shigeo Kure: kure@med.tohoku.ac.jp

Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp

Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp

Satoshi Hiyama: hiyamas@nttdocomo.com

*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

*Corresponding authors:

Junichi Sugawara: jsugawara@med.tohoku.ac.jp

and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp

Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, 980-8573, Sendai, Japan. Phone: +81-22-273-6283

Word count: 5305 words

Key words: lifelog, multi-omics analysis, prediction, complicated pregnancy

Abstract

Purpose: A prospective cohort study for pregnant women, the Maternity Log study (MLOG), was designed to construct a time-course high-resolution reference catalog 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 labor, using continuous lifestyle monitoring combined with multi-omics 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 ≥ 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 physiologic 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 multi-omics analysis.

Future plans:

Lifelog and multi-omics data will be used to construct a time-course high-resolution reference catalog of pregnancy. The reference catalog will allow us to discover relationships among multi-dimensional phenotypes and novel risk markers in pregnancy for the future personalized 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.
- Multi-omics 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 [1-4]. These multifactorial conditions are caused by an interaction of genetic factors and environmental factors [5,6]. Recent reports suggest that continuous lifestyle monitoring using wearable biosensors provides important information on latent physiologic changes that are exhibited prior to the onset of disease [7]. 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 physiologic data using multiple home healthcare devices. In addition, a variety of biospecimens are collected for multi-omics analysis.

To the best of our knowledge, this study will be the first to integrate multi-omics 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 multi-dimensional phenotypes, and lead to identification of novel risk markers in pregnancy for the future personalized early prediction of pregnancy complications, e.g. hypertensive disorders of pregnancy, gestational diabetes, and preterm labor.

COHORT DESCRIPTION

Study setting

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

The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-generation cohort study established by the Tohoku Medical Megabank Organization (TMM BirThree Cohort Study) [8] 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 Tohoku Medical Megabank Organization (ToMMo) [8]. 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, a total of over 70,000 participants [8]. Written informed consent was obtained from all participants by the genome medical research coordinators (GMRCs). The MLOG study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committees of Graduate School of Medicine (2014-1-704) and ToMMo (22017-1-085), Tohoku University under a collaborative research agreement among ToMMo, Tohoku University

and NTT DoCoMo, Inc. (Tokyo, Japan).

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 flowchart 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 ≥ 20 years and the ability to access the internet using a smartphone in the Japanese language. Participants were excluded after enrollment 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 to 35 weeks, and every week from 36 weeks to delivery [9]. 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 3 times from each participant; the first sample was collected between 12 and 24 weeks of gestation, the second between 24 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°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, NJ, USA) at -80°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°C, and in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) for separation of plasma to be stored at -80°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°C until the time of analysis.

Saliva and dental plaque sampling

Samples of saliva and dental plaque were collected 3 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, NY, USA) and stored at -80°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°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, *i.e.*, body temperature [10], home blood pressure [11], body weight [12], and physical activity (calorie expenditure) [13], as well as self-administered information such as sleep quality [14], condition of stool [15], severity of nausea [16], fetal movement [17], severity of pain [18], uterine contractions [19], and palpitations [20]. 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 [21-23]; severity of nausea using the Pregnancy-Unique Quantification of Emesis and nausea (PUQE) score [24,25]; 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 [26,27].

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 diarrhea [21, 22], and its use has been spread widely to evaluate functional bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types according to cohesion and surface cracking [21, 22].

The PUQE score [24, 25] 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 h). The total score ranges from 3 (no symptoms) to 15, and higher scores are correlated with increasing severity of nausea and vomiting [24, 25].

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 [26]. 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 9 movements [27]. 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 favorable 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 Co., Ltd., 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 meter (OMRON Healthcare Co., Ltd.) 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 Co., Ltd.) just after awakening. Physical activity was assessed using an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count steps and calculate calorie expenditure.

Clinical and epidemiological information

Baseline clinical information and maternal and neonatal outcomes (*e.g.*, 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 [8].

Database

A customized laboratory information management system (LIMS) was established to track all biospecimens. All data were transferred to the TMM integrated database after 2-step anonymisation in a linkable fashion.

Data handling was strictly regulated under HIPAA (Health Insurance Portability and Accountability Act of 1996, United States Security and Privacy Rules) [28, 29] and the Act on the Protection of Personal Information [30]. Security control at our facility has been described previously [31].

Omics analysis

Whole-genome sequencing

To minimize amplification bias, we adopted a PCR-free library preparation method. After performing library quality control using the quantitative MiSeq method [32], libraries were sequenced on HiSeq 2500 Sequencing System (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We generated the sequencing data at over 12.5x coverage on average, and we identified variants using the alignment tool BWA-MEM (ver. 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 (ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's Variant Quality Score Recalibration (VQSR) 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.megabank.tohoku.ac.jp) [33]. 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°C, total RNA was purified with PAXgene Blood RNA Kit® (Qiagen, Venlo, The Netherlands) using QiaSymphony® (Qiagen). Total RNA was reverse-transcribed using an oligo-dT primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.) 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, CA, 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 thirty millions. For computing a series of quality control metrics for RNA-seq data, RNA-SeQC was used to check the quality of sequence reads [34].

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 Corp., Billerica, MA, USA) [35]. Standard 1-dimensional nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-Meiboom-Gill (CPMG) 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 analyzed using the TopSpin 3.5 (Bruker Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta, Canada) programs. 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" [36].

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

Sample preparation for plasma and urine (50 µL each) was performed using a Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the methods previously reported by Nishiumi [37, 38]. The resulting deproteinized and derivatized supernatant (1 µL) was subjected to GC-MS/MS, performed on a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound separation was performed using a fused silica capillary column (BPX-5; 30 m × 0.25 mm inner diameter; film thickness, 0.25 µm; Shimadzu Corp, Kyoto, Japan). Metabolite detection was performed using Smart Metabolites Database (Shimadzu Corp.) 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 analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan). Then, two types of normalization were performed to these annotated metabolites. The first normalization 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 normalization was performed using quality control (QC) samples which were injected after every 12 study samples according to the RQC normalization methods [39]. Normalized 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 [40]. In brief, saliva was collected in a 50-mL tube. Dental plaque was sampled by participants by brushing teeth with a sterilized toothbrush, and then suspending it in 0.5 mL Tris-EDTA for collection. Both samples were stored at -80°C until the time of processing. DNA was extracted from saliva and dental plaque by standard glass bead-based homogenization and subsequent purification with a silica-membrane spin-column using PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin column with 30-μL RNase-free water (Takara Bio, Inc., Shiga, Japan), and stored at -20°C after determining the amount and purity of DNA with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, 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 2-step PCR. Tag-indexed PCR products thus obtained were subjected to multiplex amplicon sequencing using MiSeq System with MiSeq Sequencing Reagent Kit, v3 (Illumina, Inc.) 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 to 13 weeks of gestation using transvaginal ultrasound. HDP was defined as gestational hypertension, preeclampsia, superimposed preeclampsia, or chronic hypertension [41,42]. Preterm birth was defined as spontaneous preterm labor, medically induced preterm labor, 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 (IADPSG) criteria [43]. 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-mmHg difference in average home blood pressure (with a 7-mmHg standard deviation) 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 [44]. Vector autoregressive (VAR) modeling 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 [45]. 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 - e.g. Amelia library [46], after normalising the data by data transformation if required. For downstream

analysis, the data might be collapsed with time scale, e.g. taking trimmed mean or median for each week, month, or trimester.

Statistical analysis of multi-omics data

The present study allows combination of longitudinal lifelog data with multi-omics data. In contrast to single omics analysis, the multi-omics analysis would reveal the complicated interactions between one and another. However, the sample size for multi-omics 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 [47] and sparse factor analysis [48] 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 [49]. Several unsupervised clustering methods [50,51,52] would be also applicable to obtain hidden patterns from dense time-course lifelog measurements, which might be related to pregnancy complications. Recently developed multi-view factor analysis approaches [53,54] 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 characterize 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 analyzed using statistical hypothesis tests such as Welch's t-test, Fisher's exact test, the Chi-square test, and others as appropriate. Multiple logistic regression modeling 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 regularized algorithms (*e.g.,* LASSO, ridge regression, or elastic net) will be used to select the optimal number of contributing features that maximize 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, *e.g.*, PRSice [55], 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 6 million 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 body mass indices (BMI ≥ 25 kg/m^2). 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 were 17.0 ± 5.0 and 27.5 ± 2.5, respectively. The third blood sampling was performed at 31.1 ± 3.0 days after delivery on average. The length of enrollment ranged from 90 to 396 days with a mean of 216 ± 61 days.

Table 1. Participant characteristics

Characteristics	Value
Maternal (n = 285)	
· Age at delivery, y, mean (SD)	33.3 (± 4.9)
· Age at delivery, y, n (%)	
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)
· Education (n=81) n (%)	
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
· Occupation (n=270) n (%)	
Housewife or unemployed	93 (34.4)
Employed	175 (64.8)
Student	2 (0.7)
· Annual household income, yen (n=248) n (%)	
< 2 million	17 (6.9)
2-4 million	59 (23.8)
4-6 million	73 (29.4)
6-8 million	51 (20.6)
8-10 million	22 (8.9)
> 10 million	26 (10.5)
· Parity, n (%)	
0	140 (49.1)
1	93 (32.6)
≥ 2	52 (18.2)
· Prepregnancy BMI*, kg/m^2, mean (SD)	22.7 (± 5.1)
· Prepregnancy BMI, kg/m^2, n (%)	
< 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)
· Mode of delivery, n (%)	
Noncesarean	179 (62.8)
Cesarean	106 (37.2)
· Pregnancy complication, n (%)	
Hypertensive disorder of pregnancy	24 (8.4)

Spontaneous preterm birth	16 (5.6)
Neonatal (n = 300)	
· Birth weight, g, mean (SD)	2907 (± 572)
· Sex, n (%)	
Male	168 (56)
Female	132 (44)
· Low-birth weight (< 2500 g), n (%)	54 (18)

*BMI, body mass index

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 enrollment 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 6 million.

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 6 million total data points. Biospecimens for multi-omics 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 [8]. Second, we have successfully collected longitudinal, continuous, individual lifelog data with a high acquisition rate, which will enable us to assess dynamic changes in physiologic conditions throughout pregnancy. Third, multi-omics data will make it possible to fully understand the complex mechanisms of multifactorial pregnancy-related diseases and to overcome the unpredictability of these complications.

Prediction models using clinical and epidemiological information and circulating factors for pregnancy-related diseases have been developed extensively [56], and risk-assessment approaches using clinical information have also been developed [57, 58]. However, there is a lack of evidence for the benefits of these predictive models for routine clinical use [59]. 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 [60]. 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 multi-omics data were reported. One of the studies analyzed lifelog and multi-omics data, collected from 108 individuals at three time points during a nine-month period [61]. In their study, several remarkable relationships were identified among physiological and multi-omics 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 [49]. They studied both cis-eQTL (within 1 Mb 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 5,862 cis-eQTL and one trans-eQTL associations. These previous studies indicate that our time-course high-resolution reference catalog 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 multi-omics and lifelog data.

The MLOG study will enable us to construct a time-course high-resolution reference catalog of wellness and multi-omics data from pregnant women and thereby develop a personalized predictive model for pregnancy complications. Progressive data sharing and collaborative studies would make it possible to establish a standardized 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 [8] 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

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, Japan.

3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536.

4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai, Miyagi 981-8558, Japan.

5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan.

6 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8574, 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, Mika Hotta, Naomi Matsukawa, Makiko Sumii, Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.

Contributors

JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH, MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM, OT, SKu: recruitment and sample collection. DO, RY, TY, DS, TO, YT, YH, TFS, TM, JK, FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data processing, and statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK, NY, MY, SH, 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 (AMED) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Competing interests

This study was funded by NTT DoCoMo, Inc.

Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT DoCoMo, Inc. All other authors declare that they have no competing interests.

Ethics approval and consent to participate

TMM BirThree Cohort Study was approved by the ethics committees of the Tohoku University (authorization 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 or Dr. Masao Nagasaki at nagasaki@megabank.tohoku.ac.jp. Currently, no additional data are available.

REFERENCES

1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. Diabetes Care. 2007;30 Suppl 2: S141-6. doi: 10.2337/dc07-s206.

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. doi: 10.2471/BLT.08.062554.

3. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol. 2009; 33:130-7. doi: 10.1053/j.semperi.2009.02.010.

4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States, 1980-2010: age-period-cohort analysis. BMJ. 2013;347: f 6564. doi: 10.1136/bmj.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. doi: 10.1007/s11906-017-0718-1.

6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia / eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:

Elsevier; 2990: 51-72.

7. Li X, Dunn J, Salins D, Zhou G, et al. Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health Related Information. PLoS Biol. 2017; 15: e2001402. doi: 10.1371/journal.pbio.2001402.

8. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank Project: Design and Mission. J Epidemiol. 2016; 26:493-511. doi: 10.2188/jea.JE20150268.

9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo, Japan, pp. 1–4, 2017 [in Japanese].

10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the thermoneutral zone: a longitudinal study of normal human pregnancy. Acta Physiol (Oxf). 2011; 201: 467-74. doi: 10.1111/j.1748-1716.2010.02228.x.

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. doi: 10.1097/HJH.0b013e32831364a7.

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 Jun 11; 14: 201. doi: 10.1186/1471-2393-14-201.

13. Sorensen TK, Williams MA, Lee IM, et al. Recreational physical activity during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-80. doi: 10.1161/01.HYP.0000072270.82815.91

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. doi: 10.2337/dc11-0780.

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. doi: 10.1136/gut.2006.108324.

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. doi: 10.3109/00016340903348375.

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. doi: 10.1016/j.ejogrb.2004.12.020.

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. doi: 10.1056/NEJMoa002868

20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular system. Int J Cardiol. 2005; 98: 179-89. doi: 10.1016/j.ijcard.2003.10.028.

21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. Scand J Gastroenterol. 1997; 32: 920-4.
doi: 10.3109/00365529709011203

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. doi: 10.1007/s101510100019

23. Longstreth GF, Thompson WG, Chey WD, et al. Functional Bowel Disorders. Gastroenterology 2006; 130: 1480–91. doi: 10.1053/j.gastro.2005.11.061

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. doi: 10.1080/01443610500060651

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. doi: 10.1111/j.1471-0528.2011.02993.x

28. 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.

29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-191. US Statut Large. 1996; 110:1936-2103.

30. Amended Act on the Protection of Personal Information. 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. doi: 10.1186/s12911-017-0494-5

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. doi: 10.1016/j.ab.2014.08.015

33.Yamaguchi-Kabata Y, et al, Nariai N, Kawai Y, et al. iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. Hum Genome Var. 2015; 2:15050. doi: 10.1038/hgv.2015.50. https://www.ncbi.nlm.nih.gov/pubmed/27081555.

34. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012; 28: 1530-1532. doi: 10.1093/bioinformatics/bts196.

35. Koshiba S, Motoike I, Kojima K, et al. The structural origin of metabolic quantitative diversity. Sci Rep. 2016; 6: 31463. doi: 10.1038/srep31463

36. Vu TN, Valkenborg D, Smets K, et al. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. BMC Bioinformatics. 2011; 12: 405. doi: 10.1186/1471-2105-12-405.

37. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based

diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459. doi: 10.1371/journal.pone.0040459.

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, 17115-17126. doi: 10.18632/oncotarget.15081.

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(8): e0160555. doi: 10.1371/journal.pone.0160555.

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. doi: 10.1371/journal.pone.0131607.

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. doi: 10.1161/HYPERTENSIONAHA.117.10803.

42. Watanabe K, Naruse K, Tanaka K, et al. Outline of definition and classification of pregnancy induced hypertension (PIH). Hypertens Res Pregnancy 2013; 1: 3–4.

43. IADPSG Consensus Panel: International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682. doi: 10.2337/dc09-1848.

44. Box GEP, Jenkins GM, Reinsel GC. Time series Analysis: Forecasting and Control. 5th ed. 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, 45 (7) 2011.

47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007; 3: 1724-35. doi: 10.1371/journal.pgen.0030161

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. doi: 10.1038/nprot.2011.457.

49. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene expression across human tissues. Nature. 2017; 550: 204-213. doi:

10.1038/nature24277.

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. doi: 10.1086/660858.

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. doi: 10.1371/journal.pcbi.1005896.

52. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of Structured Time-Series. IEEE Trans Pattern Anal Mach Intell. 2015; 37: 383-93. doi: 10.1109/TPAMI.2014.2318711.

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. doi: 10.1371/journal.pcbi.1005752.

54. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018; 14: e8124. doi: 10.15252/msb.20178124.

55. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics, 2015; 31:1466-8. doi: 10.1093/bioinformatics/btu848.

912 56. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for
913 the Risk of Preterm Labor: An Update. Clin Lab Med. 2016; 36: 369-83. doi:
914 10.1016/j.cll.2016.01.019.
915
916 57. Al-Rubaie Z, Askie LM, Ray JG, et al. The performance of risk prediction
917 models for pre-eclampsia using routinely collected maternal characteristics and
918 comparison with models that include specialised tests and with clinical guideline
919 decision rules: a systematic review. BJOG. 2016; 123:1441-1452. doi:
920 10.1111/1471-0528.
921
922 58. Koullali B, Oudijk MA, Nijman TA, et al. Risk assessment and management
923 to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21: 80-8. doi:
924 10.1016/j.siny.2016.01.005.
925
926 59. Henderson JT, Thompson JH, Burda BU, et al. Preeclampsia Screening:
927 Evidence Report and Systematic Review for the US Preventive Services Task
928 Force. JAMA. 2017; 317: 1668-1683. doi: 10.1001/jama.2016.18315.
929
930 60. Broekhuijsen K, van Baaren GJ, van Pampus MG, et al; HYPITAT-II Study
931 Group. Immediate delivery versus expectant monitoring for hypertensive
932 disorders of pregnancy between 34 and 37 weeks of gestation (HYPITAT-II): an
933 open-label, randomised controlled trial. Lancet. 2015; 385: 2492-2501. doi:
934 10.1016/S0140-6736(14)61998-X.
935
936 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: 747-756. doi: 10.1038/nbt.3870.

FIGURE TITLES AND LEGENDS

Figure 1. Flowchart of Maternity Log Study (MLOG) participants

Figure 2. Overview of the MLOG study protocol

A: Participant timeline for the MLOG study.

B: Physiologic information collected using healthcare devices. Specific measures were uploaded each day from the time of enrollment (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 enrollment (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 and 20 weeks of gestation, respectively.

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 enrollment to delivery for each participant.

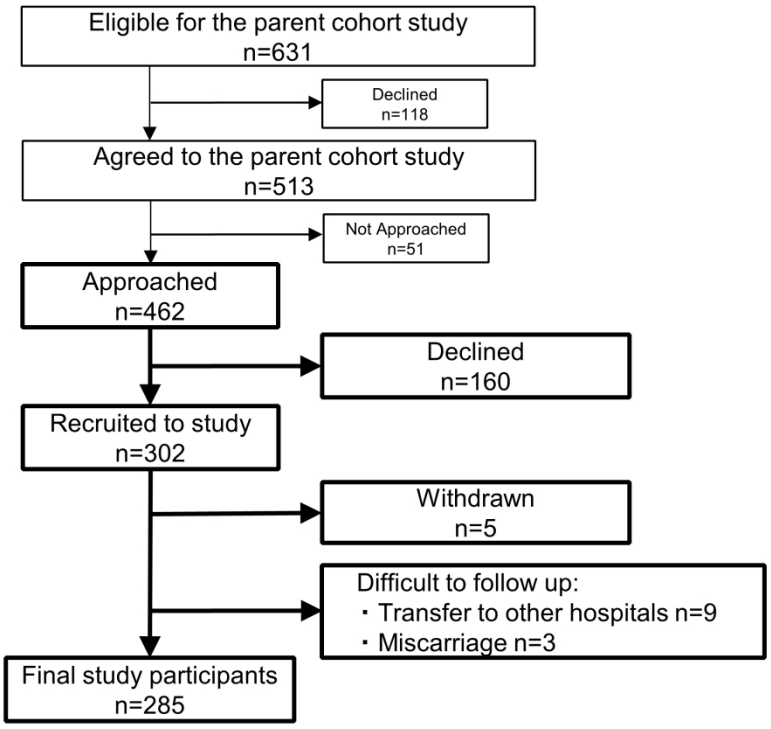

Figure 1.

254x190mm (300 x 300 DPI)

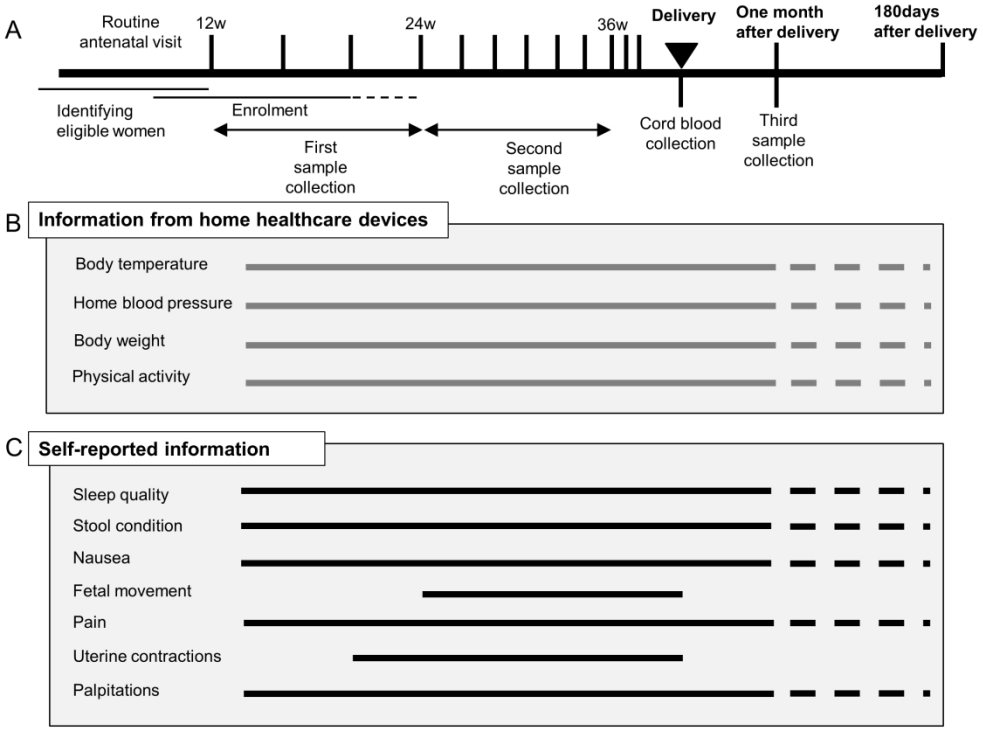

Figure 2.

254x190mm (300 x 300 DPI)

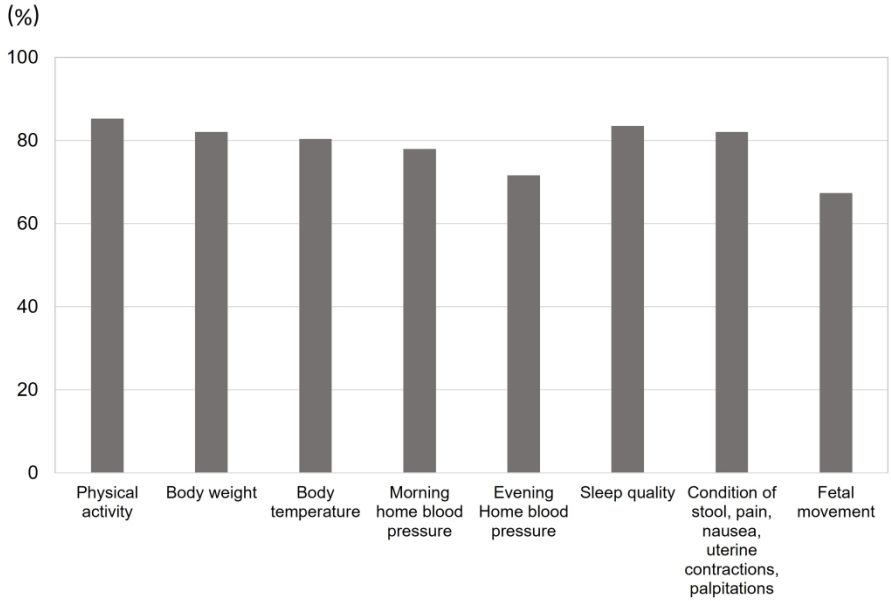

Figure 3.

254x190mm (300 x 300 DPI)