Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<u>http://bmjopen.bmj.com</u>).

If you have any questions on BMJ Open's open peer review process please email <u>editorial.bmjopen@bmj.com</u>

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; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; 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 Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical Megabank Organization, Tohoku University Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University Kushiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University Kushiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University; Tohoku University Graduate School of Medicine Kinoshita, Kengo; Tohoku Medical Megabank Organization, Tohoku University

1 2 3	
4 5 6 7	
8 9 10	
11 12 13 14	
15 16 17	
18 19 20 21	
22 23 24	
25 26 27 28	
29 30 31 32	
33 34 35	
36 37 38 39	
40 41 42	
43 44 45 46	
47 48 49	
50 51 52 53	
54 55 56	

	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
	SCHOLARONE [™] Manuscripts
	SCHOLARONEM Manuscripts
F	only http://bmienen.hmi.com/site/about/guidelines.yhtml

of for perv first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . gg Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1	Maternity Log Study: protocol for a longitudinal lifelog
2	monitoring and multi-omics analysis for the early
3	prediction of complicated pregnancy
4	
5	Junichi Sugawara ^{1,2,3} *, Daisuke Ochi ^{1,4} , Riu Yamashita ¹ , Takafumi Yamauchi ^{1,4} ,
6	Daisuke Saigusa ¹ , Maiko Wagata ^{1,2} , Mami Ishikuro ¹ , Yoshiki Tsunemoto ⁴ , Yuki
7	Harada ¹ , Tomoko F.Shibata ¹ , Junko Kawashima ¹ , Fumiki Katsuoka ¹ , Takako
8	Igarashi-Takai ¹ , Soichi Ogishima ¹ , Hirohito Metoki ⁵ , Hiroaki Hashizume ¹ , Nobuo
9	Fuse ^{1,2} , Naoko Minegishi ¹ , Seizo Koshiba ¹ , Osamu Tanabe ¹ , Shinichi
10	Kuriyama ^{1,2} , Kengo Kinoshita ¹ , Shigeo Kure ^{1,2} , Nobuo Yaegashi ^{1,2,3} , Masayuki
11	Yamamoto ^{1,2} , Satoshi Hiyama ⁴ , and Masao Nagasaki ¹ *.
12	
13	1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-
14	machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.
15	2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku,
16	Sendai, Miyagi, 980-8574, Japan.
17	3 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-
18	8574, Japan.
19	4 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,
20	Kanagawa, Japan 239-8536.
21	5 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima,
22	Aobaku, Sendai, Miyagi 981-8558, Japan.
23	

Page 3 of 35

1

2		
3	24	Email addresses:
5 6	25	*Junichi Sugawara: jsugawara@med.tohoku.ac.jp
7 8	26	Daisuke Ochi: ochi@megabank.tohoku.ac.jp
9 10	27	Riu Yamashita: ryamasi@megabank.tohoku.ac.jp
11 12	28	Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp
13 14 15	29	Daisuke Saigusa: saigusa@m.tohoku.ac.jp
16 17	30	Maiko Wagata: wagata@med.tohoku.ac.jp
18 19	31	Mami Ishikuro: m_ishikuro@med.tohoku.ac.jp
20 21	32	Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp
22 23	33	Yuki Harada: harada@megabank.tohoku.ac.jp
24 25	34	Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp
26 27	35	Junko Kawashima: kawashima@dent.tohoku.ac.jp
28 29 30	36	Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp
31 32	37	Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp
33 34	38	Soichi Ogishima: ogishima@megabank.tohoku.ac.jp
35 36	39	Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp
37 38	40	Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp
39 40	41	Nobuo Fuse: fusen@megabank.tohoku.ac.jp
41 42	42	Naoko Minegishi: nmine@med.tohoku.ac.jp
43 44	43	Seizo Koshiba: koshiba@megabank.tohoku.ac.jp
45 46 47	44	Osamu Tanabe: otanabe@megabank.tohoku.ac.jp
48 49	45	Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp
50 51	46	Kengo Kinoshita: kengo@ecei.tohoku.ac.jp
52 53	47	Shigeo Kure: kure@med.tohoku.ac.jp
54 55	48	Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp
56 57		
58 59		For peer review only - http://bmiopen.hmi.com/site/about/guideling

pert-first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . g Pa

BMJ	Open
-----	------

49	Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp
50	Satoshi Hiyama: hiyamas@nttdocomo.com
51	*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp
52	
53	*Corresponding authors:
54	Junichi Sugawara: jsugawara@med.tohoku.ac.jp
55	and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp
56	Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi,
57	Aoba-ku, 9808573, Sendai, Japan. Phone: +81-22-273-6283
58	
59	Word count: 3931 words
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	

Page 5 of 35

1	
2	
3 4	
5	
6	
6 7	
8	
9	
10	
11 12	
13	
14	
15	
12 13 14 15 16 17	
17	
18 19	
20	
21	
22	
23	
19 20 21 22 23 24 25	
26	
27	
28	
29	
30 21	
31 32	
33	
34	
35	
36 37	
37	
39	
40	
41	
42	
43 44	
45	
46	
47	
48	
49 50	
51	
52	
53	
54	
55 56	
50 57	
58	
59	
60	

72 Abstract

Introduction: Multifactorial diseases, including various complications of 73 74 pregnancy, are caused by a complex interaction of genetic and environmental 75 factors such as lifestyle and living environment. The evaluation of continuous 76 lifestyle monitoring using healthcare devices provides information on latent 77 physiologic changes prior to the onset of disease. We expect that monitoring 78 these factors directly is more accurate than using conventional methods such 79 as questionnaires. A prospective cohort study for pregnant women, the 80 Maternity Log study (MLOG), was designed to construct a time-course highresolution reference catalog of bioinformatic data in pregnancy and explore the 81 82 associations between genomic and environmental factors and the onset of 83 pregnancy complications using continuous lifestyle monitoring combined with 84 multi-omics data on the genome, transcriptome, proteome, metabolome, and 85 microbiome. Methods and analysis: Pregnant women were recruited at the timing of first 86 routine antenatal visits. Study participants uploaded daily general health 87

88 information including quality of sleep, condition of bowel movements, and the 89 presence of nausea, pain, and uterine contractions. Participants also collected 90 physiologic data, such as body weight, blood pressure, heart rate, and body 91 temperature, using multiple home healthcare devices. Biospecimens, including 92 maternal plasma, serum, urine, saliva, dental plaque, and cord blood, were collected for multi-omics analysis. This study is expected to elucidate the causal 93

- 94 relationship between complicated pregnancy and maternal lifestyle and
- 95 physiologic changes. Lifelog and multi-omics data will be used to construct a
- 96 time-course high-resolution reference catalog of pregnancy. The reference

pen: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Superieur (ABES)

35 of first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Page Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3	
4 5	
6	
7	
8 9	
10	
11 12	
13	
14 15	
16	
17	
10	
20	
21	
23	
24 25	
26	
27	
29	
4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39	
32	
33 34	
35	
36 37	
38	
39 40	
41	
42 43	
44	
45 46	
47	
48 49	
50	
51 52	
52 53	
54 55	
55 56	
57	
58 59	
60	

1 2

- 98 phenotypes and novel risk markers in pregnancy for the future personalized
- 99 early prediction of pregnancy complications.
- 100 Ethics and dissemination: This study was approved by the Tohoku Medical
- 101 Megabank Organization, Tohoku University (2014-1-704 and 2017-1-085).
- 102 Written informed consent was obtained from all participants.

103

- 104 Strengths and limitations of this study:
- 105 This is the first study designed to collect longitudinal lifelog information through
- 106 healthcare devices, self-administered questionnaires using smartphones, and
- 107 varieties of biospecimens throughout pregnancy.
- 108 Longitudinal, continuous, individual lifelog data with a high acquisition rate will
- 109 enable us to assess dynamic physiological changes throughout pregnancy.
- 110 Mutli-omics data will make it possible to understand the complex mechanisms
- 111 of multifactorial pregnancy-related diseases.
- 112 A time-course high-resolution reference catalog of wellness and multi-omics
- 113 data will be informative to develop a personalized predictive model for
- 114 pregnancy complications.
- 115 Further study with larger sample size is needed to validate a reference catalog
- of normal pregnancy and a prediction model of pregnancy complications.
- 117
- 118

120

- 119
- - 121

Page 7 of 35

122 INTRODUCTION

The incidence of pregnancy-related disorders, including hypertensive disorders of pregnancy (HDP), gestational diabetes mellitus (GDM), and preterm deliverv 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 guestionnaires. 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.

141 METHODS AND ANALYSIS

142 Study setting

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

pen: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
33 34	
34 35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
52	
55 54	
54 55	
56	
57	
58	
59	
60	

1 2 3

147	transcriptomes, metabolomes, and oral microbiomes.
148	The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-
149	generation cohort study established by the Tohoku Medical Megabank
150	Organization (TMM BirThree Cohort Study) [8] in order to elucidate the
151	mechanisms of complicated multifactorial diseases in mothers and children in
152	the wake of the Great East Japan Earthquake in 2011. Epidemiological data
153	from extensive questionnaire surveys and accurate clinical records, including
154	birth outcomes, can be abstracted from the integrated biobank of Tohoku
155	Medical Megabank Organization (ToMMo) [8].
156	Written informed consent was obtained from all participants by the genome
157	medical research coordinators (GMRCs). This study was conducted in
158	accordance with the Declaration of Helsinki and approved by the ethics
159	committee of ToMMo, Tohoku University (2014-1-704, 2017-1-085). This study
160	was conducted under a collaborative research agreement with ToMMo, Tohoku
161	University and NTT DoCoMo, Inc. (Tokyo, Japan).

162

- Patient and public involvement 163
- 164 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 165
- 166 available in the public domain.

167

Participants 168

- Participants were recruited at a first routine antenatal visit at Tohoku University 169
- 170 Hospital, Sendai, Japan between September 2015 and September 2016. A
- flowchart of the recruitment process is shown in Figure 1. Patients who already 171

172	agreed to participate in the TMM BirThree Cohort Study were recruited to
173	provide an additional informed consent for the MLOG study. A total of 302
174	women were enrolled. The inclusion criteria were age \geq 20 years and the ability
175	to access the internet using a smartphone in the Japanese language.
176	Participants were excluded after enrollment if termination of pregnancy,
177	abortion, or transfer to another institution for emergency care occurred before
178	delivery, or if they withdrew consent for any reason.
179	
180	Outline of study protocol
181	The study protocol consisted of blood and urine sampling, saliva and dental
182	plaque sampling, self-administered daily lifelog data collection, and data upload
183	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

pen: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement

Superiour (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Superieur (ABES)

197	stored in a PAXgene® tube (Becton, Dickinson and Company, Franklin Lakes,	
198	NJ, USA) at -80°C until the time of RNA extraction for transcriptome analysis.	
199	Genomic DNA was extracted from mononuclear cells using an Autopure $\ensuremath{\mathbb{R}}$	
200	extractor (Qiagen, Venlo, The Netherlands). Approximately 10 mL of blood was	6
201	collected from the umbilical vein in a PAXgene® tube for storage at -80°C, and	
202	in an EDTA 2K tube (Becton, Dickinson and Company, Franklin Lakes, NJ,	
203	USA) for separation of plasma to be stored at -80°C. Urine samples (10 mL)	
204	were collected at each antenatal visit; when participants were admitted to the	
205	hospital ward, urine was collected once weekly. Urine samples were	
206	immediately transferred and stored at -80°C until the time of analysis.	
207		
208	Saliva and dental plaque sampling	
209	Samples of saliva and dental plaque were collected 3 times from each	
210	participant, at the same time points as blood collection. Approximately 3 mL of	
211	saliva was collected using a 50-mL conical centrifuge tube (Corning, Inc.,	
212	Corning, NY, USA) and stored at -80°C until analysis. Dental plaque was	
213	sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM	
214	EDTA; pH, 8.0), and immediately stored at -80°C until the time of sample	
215	processing.	
216		
217	Lifelog data collection	
218	Based on previous publications on the utility for risk assessment of pregnancy-	
219	related diseases, we selected several lifelog parameters to employ in this study	/,
220	<i>i.e.</i> , body temperature [10], home blood pressure [11], body weight [12],	
221	physical activity (calorie expenditure) [13], as well as self-administered	
		9

BMJ Open

1 2		
3 4	222	information such as sleep quality [14], condition of stool [15], severity of nausea
5 6	223	[16], fetal movement [17], severity of pain [18], uterine contractions [19], and
7 8	224	palpitations [20]. Body temperature, home blood pressure, body weight, and
9 10	225	physical activity were uploaded from multiple healthcare devices through a
11 12 12	226	smartphone. The self-administered information described above was input
13 14 15	227	manually on mobile applications created for this study.
16 17	228	Data collection was started after obtaining informed consent and after giving
18 19	229	detailed instructions for the use of the healthcare devices. These applications
20 21	230	tracked quality of sleep; condition of stool using the Bristol Scale [21-23];
22 23	231	severity of nausea using the Pregnancy-Unique Quantification of Emesis and
24 25	232	nausea (PUQE) score [24,25]; headache, toothache, lumbago, and upper and
26 27 28	233	lower abdominal pain using a numerical rating scale (NRS) score; the number
28 29 30	234	of perceived uterine contractions; palpitations; and fetal movement using a
31 32	235	modified count-to-10 fetal movement chart [26,27].
33 34	236	Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction
35 36	237	(ranked from satisfied to poor using a numeric scale of 0-4), and the number of
37 38	238	nocturnal awakenings (0-6).
39 40	239	The Bristol stool form scale was originally developed to assess constipation
41 42 43	240	and diarrhea [21, 22], and its use has been spread widely to evaluate functional
45 44 45	241	bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types
46 47	242	according to cohesion and surface cracking [21, 22].
48 49	243	The PUQE score [24, 25] was developed to estimate the severity of nausea
50 51	244	and vomiting in pregnancy and quantifies the number of daily vomiting and
52 53	245	retching episodes and the length of nausea in hours (over the preceding 12 h).
54 55	246	The total score ranges from 3 (no symptoms) to 15, and higher scores are
56 57 58 59		10

pen: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

3
4
5
6
7
8
9
10
11
12
13
14
15
16 17
17
18
19
20
21
77
22
24
25
26 27
27 28
28 29
30
31
32
33
34
35
36 37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54 55
55 56
56 57
57 58
58 59
59 60
00

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).

251 Uterine contractions and palpitations were evaluated using definitions 252 determined for the current study. Uterine contractions were assessed using the 253 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 254 255 recording the time, in minutes, required to count 10 fetal movements [26]. More 256 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 257 time required to perceive an additional 9 movements [27]. Pregnant women are 258 259 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 260 261 maternal position after 24 weeks of gestation. 262 The applications also collected dietary logs and the medications taken on the 263 day before and the day of the antenatal visit, on which blood or urine samples 264 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

1		
2		
3		
4		
5		
6		
7		
8 9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19 20		
20		
21 22		
23		
24		
25		
26		
27		
28		
29		
30		
31 32		
32 33		
33 34		
35		
36		
37		
38		
39		
40		
41		
42		
43 44		
44 45		
45 46		
47		
48		
49		
50		
51		
52		
53		
54		
55 56		
56 57		
57 58		
50 59		
60		

272	daily within 1 hour of awakening in the morning. Daily body temperature was	
273	evaluated using an MC-652LC digital thermometer (MC-652LC; OMRON	
274	Healthcare Co., Ltd.) just after awakening. Physical activity was assessed using	ļ
275	an HJA-403C pedometer (HJA-403C; OMRON Healthcare Co., Ltd.) to count	
276	steps and calculate calorie expenditure.	
277		
278	Clinical and epidemiological information	
279	Baseline clinical information and maternal and neonatal outcomes (e.g.,	
280	maternal age, clinical data and findings from each antenatal visit, gestational	
281	age at delivery, type of delivery, birth weight, maternal and fetal complications)	
282	were obtained from the medical records of the Tohoku University Hospital.	
283	Epidemiological data, including extensive questionnaire surveys from the TMM	
284	BirThree Cohort can be obtained from the ToMMo integrated biobank [8].	
285		
286	Database	
287	A customized laboratory information management system (LIMS) was	
288	established to track all biospecimens. All data were transferred to the TMM	
289	integrated database after 2-step anonymization in a linkable fashion.	
290	Data handling was strictly regulated under HIPAA (Health Insurance Portability	
291	and Accountability Act of 1996, United States Security and Privacy Rules) [28,	
292	29] and the Act on the Protection of Personal Information [30]. Security control	
293	at our facility has been described previously [31].	
294		
295	Omics analysis	
296	Whole-genome sequencing	
	12	2

1 2 3 4 5 6	
7 8 9 10 11 12	
13 14 15 16 17 18	
19 20 21 22 23 24 25	
25 26 27 28 29 30 31	
32 33 34 35 36 37	
38 39 40 41 42 43	
44 45 46 47 48 49	
50 51 52 53 54 55	
56 57 58 59 60	

297	To minimize amplification bias, we adopted a PCR-free library preparation	
298	method. After performing library quality control using the quantitative MiSeq	
299	method [32], libraries were sequenced on HiSeq 2500 Sequencing System	
300	(Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We	Э
301	generated the sequencing data at over 12.5x coverage on average, and we	
302	identified variants using the alignment tool BWA-MEM (ver. 0.7.5a-r405) with	
303	the default option. Single nucleotide variants (SNVs) and indels were jointly	
304	called across all samples using Genome Analysis Tool Kit's HaplotypeCaller	
305	(ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's	
306	Variant Quality Score Recalibration (VQSR) approach. The human reference	
307	genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC_00760)5
308	(Human Gamma Herpesvirus 4). The complete fasta file named	
309	hg19_tommo_v2.fa is available from iJGVD website	
310	(http://ijgvd.megabank.tohoku.ac.jp).	
311	Transcriptome	
312	Transcriptome	
313	Whole blood were collected using the PAXgene® RNA tube, which is widely	
314	used for transcriptome analysis. After storage at -80°C, total RNA was purified	
315	with PAXgene Blood RNA Kit $^{ m I\!R}$ (Qiagen, VenIo, The Netherlands) using	
316	QiaSymphony ${ m I\!R}$ (Qiagen). The amount and quality of the total RNA was	
317	assessed with Bio Analyzer ${\ensuremath{\mathbb R}}$ or Tape Station ${\ensuremath{\mathbb R}}$ (both from Agilent	
318	Technologies, Santa Clara, CA, USA), and we only used RNA samples with an	۱
319	RNA integrity number (RIN) (or an RIN equivalent) higher than 7.0. Total RNA	
320	was reverse-transcribed using an oligo-dT primer. We used TruSeq DNA PCR	-
321	Free Library Preparation Kit (Illumina, Inc.) for library preparation for	
		13

2		
3 4	322	sequencing with HiSeq 2500 Sequencing System.
5 6	323	
7 8	324	Plasma and urine metabolome
9 10	325	Nuclear magnetic resonance (NMR) spectroscopy
11 12	326	All NMR measurements for metabolome analysis were conducted at 298 K on a
13 14 15	327	Brucker Avance 600 MHz spectrometer equipped with a SampleJet sample
16 17	328	changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional
18 19	329	nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-
20 21	330	Meiboom-Gill (CPMG) spectra were obtained for each plasma or urine sample.
22 23	331	All spectra for plasma or urine samples were acquired using 16 scans and 32 k
24 25	332	of complex data points. All data were analyzed using the TopSpin 3.5 (Bruker
26 27	333	Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta,
28 29 30	334	Canada) programs.
31 32	335	
33 34	336	Gas chromatography-tandem mass spectrometry (GC-MS/MS)
35 36	337	Sample preparation for plasma and urine (50 µL each) was performed using a
37 38	338	Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the
39 40	339	methods previously reported by Nishiumi [36, 37]. The resulting deproteinized
41 42	340	and derivatized supernatant (1 μ L) was subjected to GC-MS/MS, performed on
43 44	341	a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound
45 46 47	342	separation was performed using a fused silica capillary column (BPX-5; 30 m ×
48 49	343	0.25 mm inner diameter; film thickness, 0.25 µm; Shimadzu Corp, Kyoto,
50 51	344	Japan). Metabolite detection was performed using Smart Metabolites Database
52 53	345	(Shimadzu Corp.) that contained the relevant multiple reaction monitoring
54 55	346	(MRM) method file and data regarding the GC analytical conditions, MRM
56 57	540	
58		די די

pend first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement 1 1 9 9 Pa

BMJ Open

347 parameters, and retention index employed for the metabolite measurement. The

348 database used in this study included data on 475 peaks from 334 metabolites.

349 All peaks of metabolites detected from each sample was annotated and

analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan).

352 Oral Microbiome

Analysis of oral microbiome was conducted by previously reported protocols [36]. In brief, saliva was collected in a 25-mL tube. Dental plague 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 guality 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 MiSeg System and MiSeg Sequencing Reagent Kit, v3 (Illumina, Inc.) according to the manufacturer's instructions. Outcomes

371 The following obstetric complications represented the primary outcomes. HDP

372	was classified as gestational hypertension, preeclampsia, superimposed
373	preeclampsia, and chronic hypertension [37]. Spontaneous preterm birth was
374	defined as spontaneous preterm labor or preterm premature rupture of
375	membranes resulting in preterm birth at less than 37 weeks of gestation. GDM
376	was diagnosed according to the International Association of the Diabetes and
377	Pregnancy Study Groups (IADPSG) criteria [38]. The secondary outcomes were
378	maternal body weight, blood pressure, physical activity, lifestyle changes,
379	perinatal mental disorders, fetal growth, fetal movement, and birth weight.
380	
381	Statistical analysis
382	The association of outcomes with each factor will be analyzed using a statistical
383	hypothesis test such as Welch's t-test, Fisher's exact test, the Chi-square test,
384	and others as appropriate. Multiple logistic regression modelling will be used to
385	adjust for confounders and to assess whether each factor or combination of
386	factors can be used to predict outcomes. Stepwise selection algorithms or
387	regularized algorithms (<i>e.g.,</i> LASSO, ridge regression, or elastic net) will be
388	used to select the optimal number of contributing factors that maximize the
389	predictive power using the leave-1-out cross validation or K-fold cross validation
390	methods.
391	Individual genetic factors may have an effect on outcomes; therefore, some
392	aggregated genetic risk score should be included in the prediction model. For
393	example, SNVs, including rare variants in or around a chromosome region of a
394	known or estimated risk gene, could be aggregated by considering their impacts
395	on biological function of the gene or their minor allele frequencies in the
396	population. However, we are limited in the number of study participants, and the
	16

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

397	aggregated risk score might therefore contril	oute only slightly to the predictive
398	power. To create a more reliable risk score,	the estimates from other large-
399	scale cohort data using polygenic score tool	s, e.g., PRSice [39], could be used
400	for this study.	
401		
402	INTERIM RESULTS	
403	Clinical background	
404	A total of 302 women were enrolled, and the	mean gestational weeks of
405	recruitment was 16.4 ± 4.9 weeks (mean ± S	SD). A total of 285 participants have
406	delivered; their baseline clinical characteristi	cs are described in Table 1. The
407	mean maternal age at delivery was 33.3 ± 4	.9 years. Approximately 42% of the
408	participants were over 35 years of age, 51%	were parous, and 22% were
409	overweight or obese by their prepregnancy b	body mass indices (BMI ≥ 25
410	kg/m ²). Overall, 8.4% of participants had HD	P, and 5.6% underwent
411	spontaneous preterm birth. On average, infa	ints were delivered at 38.0 ± 2.3
412	weeks of gestation with a mean birth weight	of 2907 \pm 572 g. The rate of low
413	birth weight was 18%. Mean gestational wee	eks of the first and second blood
414	sampling were 17.0 \pm 5.0 and 27.5 \pm 2.5, res	spectively. The third blood sampling
415	was performed at 31.1 ± 3.0 days after deliv	ery on average. The length of
416	enrollment ranged from 90 to 396 days with	a mean of 216 \pm 61 days.
417 418		
	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) 1

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

421	Data acquisition	
420		
419	*BMI, body mass index	
	 Low-birth weight (< 2500 g), n (%) 	54 18)
	Female	132 (44)
	Male	168 (56)
	• Sex, n (%)	
	• Birth weight, mean (SD)	2907 (± 572)
	Neonatal (n = 300)	
	Spontaneous preterm birth	16 (5.6)
	Pregnancy complication, n (%) Hypertensive disorder of pregnancy	24 (8.4)
		100 (07.2)
	Noncesarean Cesarean	179 (62.8) 106 (37.2)
	Mode of delivery, n (%)	170 (62 8)
	 Gestational weeks at delivery, mean (SD) 	38.0 (± 2.3)
	≥ 30.0	29 (10.2)
	25.0-29.9	34 (11.9)
	18.5-24.9	186 (65.3)
	< 18.5	36 (12.6)
	 Prepregnancy BMI*, kg/m², mean (SD) Prepregnancy BMI, kg/m², n (%) 	22.7 (± 5.1)
	≥2	52 (18.2)
	1	93 (32.6)
	• Parity, n (%) 0	140 (49.1)
	45-49	1 (0.4)
	40-44	30 (10.5)
	35-39	90 (31.6)

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

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

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

BMJ Open

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
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
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.
20
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

2	
2	
3	
4	
3 4 5 6 7 8 9 10	
6	
7	
/	
8	
9	
10	
11	
12	
12	
15	
14	
15	
16	
17	
10	
10	
12 13 14 15 16 17 18 19	
20	
21	
20 21 22 23 24 25 26 27 28 29 30 31 32 33 34	
22	
23	
24	
25	
26	
27	
28	
20	
29	
30	
31	
32	
33	
21	
34	
35	
36	
34 35 36 37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

480	Hopefully, our study will result in the development of a novel stratification
481	model for pregnancy-related diseases employing multi-omics and lifelog data.
482	The MLOG study will enable us to construct a time-course high-resolution
483	reference catalog of wellness and multi-omics data from pregnant women and
484	thereby develop a personalized predictive model for pregnancy complications.
485	Progressive data sharing and collaborative studies would make it possible to
486	establish a standardized early-prediction method through large clinical trials.
487	
488	Author affiliations
489	1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-
490	machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.
491	2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku,
492	Sendai, Miyagi, 980-8574, Japan.
493	3 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-
494	8574, Japan.
495	4 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,
496	Kanagawa, Japan 239-8536.
497	5 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima,
498	Aobaku, Sendai, Miyagi 981-8558, Japan.
499	
500	Acknowledgements
501	The authors would like to thank all the MLOG study participants, the staff of the
502	Tohoku Medical Megabank Organization, Tohoku University (a full list of
503	members is available at: http://www.megabank.tohoku.ac.jp/english/a161201/),
504	and the department of Obstetrics and Gynecology, Tohoku University Hospital,
	21

505	for their efforts and contributions. The MLOG study group also included Chika
506	Igarashi, Motoko Ishida, Yumiko Ishii, Hiroko Yamamoto, Akiko Akama, Kaori
507	Noro, Miyuki Ozawa, Yuka Narita, Junko Yusa, Miwa Meguro, Michiyo Sato,
508	Miyuki Watanabe, Mai Tomizuka, Mika Hotta, Naomi Matsukawa, Makiko Sumii,
509	Ayako Okumoto, Yukie Oguma, Ryoko Otokozawa, Toshiya Hatanaka, Sho
510	Furuhashi, Emi Shoji, Tomoe Kano, Riho Mishina, and Daisuke Inoue.
511	
512	Contributors
513	JS, DO, RY, TY, HM, OT, SKu, NY, SH, MN were involved in initial stages of
514	the strategy and design of study conception. JS, DO, RY, TY, OT, DS, SKo, SH,
515	MN: responsible for the draft of the manuscript. JS, DO, RY, TY, MW, MI, HM,
516	OY, SKu: recruitment and sample collection. DO, RY, TY, DS, YT, YH, TFS, JK,
517	FK, TIT, SO, NM, SKo, OT, MN: sample analysis, data processing, and
518	statistical analysis. JS, HH, NF, NM, SKo, OT, SKu, KK, SK, NY, MY, SH, MN:
519	advice and supervision of sample analysis. All authors have contributed to
520	revision and have approved the final manuscript.
521	
522	Funding
523	The present study was supported by NTT DoCoMo, Inc., with a collaborative
524	research agreement between NTT DoCoMo and ToMMo. This work was
525	supported in part by the Tohoku Medical Megabank Project from the Japan
526	Agency for Medical Research and Development (AMED) and the Ministry of
527	Education, Culture, Sports, Science and Technology (MEXT).
528	
529	Competing interests
	22

3	
4	
5	
6	
6 7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
20 21 22 23 24 25	
26 27	
27	
28	
29	
30	
21	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
40 41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	

530	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

BMJ Open

554	1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public
555	health perspective. Diabetes Care. 2007;30 Suppl 2: S141-6.
556	
557	2. Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The
558	worldwide incidence of preterm birth: a systematic review of maternal mortality
559	and morbidity. Bull World Health Organ. 2010; 88:31-8.
560	
561	3. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol.
562	2009; 33:130-7.
563	
564	4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United
565	States, 1980-2010: age-period-cohort analysis. BMJ. 2013;347: f 6564.
566	
	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.
570	
571	6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia /
572	eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:
573	Elsevier; 2990: 51-72.
574	
575	7. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schu "ssler-Fiorenza Rose SM, et
576	al. Digital Health: Tracking Physiomes and Activity Using Wearable
577	Biosensors Reveals Useful Health Related Information. PLoS Biol. 2017; 15:
578	e2001402.
	24
	 555 556 557 558 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 576 576 575 576 576 576 575 576 577

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

	BMJ Open	Pag
579		
580	8. Kuriyama S, Yaegashi N, Nagami F, Arai T, Kawaguchi Y, Osumi N, et al.	
581	The Tohoku Medical Megabank Project: Design and Mission.	PT
582	J Epidemiol. 2016; 26:493-511.	otecte
583		готестеа ву соругідпт,
584	9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical	орупд
585	Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo,	=
586	Japan, pp. 1–4, 2017 [in Japanese].	nciuaing
587		TOP USES
588	10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the	ses rei
589	thermoneutral zone: a longitudinal study of normal human pregnancy. Acta	ated to
590	Physiol (Oxf). 2011; 201: 467-74.	о техт а
591		and da
592	11. Metoki H, Ohkubo T, Watanabe Y, Nishimura M, Sato Y, Kawaguchi M,	ta min
593	Hara A, Hirose T, Obara T, Asayama K, Kikuya M, Yagihashi K, Matsubara Y,	ning, A
594	Okamura K, Mori S, Suzuki M, Imai Y; BOSHI Study Group. Seasonal trends of	i train
595	blood pressure during pregnancy in Japan: the babies and their parents'	raining, and
596	longitudinal observation in Suzuki Memorial Hospital in Intrauterine Period	S
597	study. J Hypertens. 2008; 26: 2406-13.	miar technologies.
598		innoio
599	12. Haugen M, Brantsæter AL, Winkvist A, Lissner L, Alexander J, Oftedal B,	gies.
600	Magnus P, Meltzer HM. Associations of pre-pregnancy body mass index and	
601	gestational weight gain with pregnancy outcome and postpartum weight	
602	retention: a prospective observational cohort study. BMC Pregnancy Childbirth.	
603	2014 Jun 11; 14: 201.	
	25	

uthy DA.
sia.
0A, Van
ce in
A meta-
v. Gut.
icental
JF.
ig with
cand.
tionship
stet
26
ml

2	
3	
4 5	
6	
4 5 7 8 9 10 11	
8	
9 10	
11	
12	
13	
14	
13 14 15 16 17	
17	
18	
19 20	
21	
22	
23	
21 22 23 24 25 26 27	
26	
27	
28 29	
30	
21	
32 33 34	
33 34	
35	
35 36 37	
37	
38 39	
40	
41	
42 43	
43 44	
45	
46	
47 48	
49	
50	
51 52	
52 53	
54	
55	
56	
57 58	
59	
60	

629	19. Iams JD, Newman RB, Thom EA, Goldenberg RL, Mueller-Heubach E,
630	Moawad A, Sibai BM, Caritis SN, Miodovnik M, Paul RH, Dombrowski MP,
631	Thurnau G, McNellis D; National Institute of Child Health and Human
632	Development Network of Maternal-Fetal Medicine Units. Frequency of uterine
633	contractions and the risk of spontaneous preterm delivery. N Engl J Med. 2002 ;
634	346: 250-5.
635	
636	20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular
637	system. Int J Cardiol. 2005; 98: 179-89.
638	
639	21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit
640	time. Scand J Gastroenterol. 1997; 32: 920-4.
641	
642	22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical
643	practice and clinical research. Tech Coloproctol 2001; 5: 163–4
644	
645	23. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller
646	RC. Functional Bowel Disorders. Gastroenterology 2006; 130: 1480–91.
647	
648	24. Koren G, Boskovic R, Hard M, Maltepe C, Navioz Y, Einarson A.
649	Motherisk-PUQE (pregnancy-unique quantification of emesis and nausea)
650	scoring system for nausea and vomiting of pregnancy. Am J Obstet Gynecol.
651	2002;186: S228-31.
652	
653	25. Koren G, Piwko C, Ahn E, Boskovic R, Maltepe C, Einarson A, et al.
	27

Page 29 of 35

1 2		
2 3 4	654	Validation studies of the Pregnancy Unique-Quantification of Emesis (PUQE)
5 6	655	scores. J Obstet Gynaecol. 2005; 25: 241-4.
7 8	656	
9 10	657	26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. Br
11 12	658	Med J 1976; 1:1305–7.
13 14	659	
15 16 17	660	27. Winje BA, Saastad E, Gunnes N, Tveit JV, Stray-Pedersen B, Flenady V, et
18 19	661	al. Analysis of 'count-to-ten' fetal movement charts: a prospective cohort
20 21	662	study. BJOG. 2011;118: 1229-38.
22 23	663	
24 25	664	28. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach
26 27	665	Notification rules under the Health Information Technology for Economic and
28 29	666	Clinical Health Act and the Genetic Information Nondiscrimination Act; other
30 31 32	667	modifications to the HIPAA rules. Fed Regist. 2013; 78: 5565-702.
33 34	668	
35 36	669	29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-
37 38	670	191. US Statut Large. 1996; 110:1936-2103.
39 40	671	
41 42	672	30. Amended Act on the Protection of Personal Information.
43 44 45	673	https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Informatio
45 46 47	674	n.pdf.
48 49	675	
50 51	676	31. Takai-Igarashi T, Kinoshita K, Nagasaki M, Ogishima S, Nakamura N,
52 53	677	Nagase S, et al. Security controls in an integrated Biobank to protect privacy in
54 55	678	data sharing: rationale and study design. BMC Med Inform Decis Mak. 2017;
56 57		28
58 59 60		For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml
00		

1

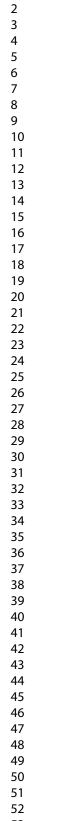
2		
3 4	679	17:100.
5	680	
6 7 8	681	32. Katsuoka F, Yokozawa J, Tsuda K, Ito S, Pan X, Nagasaki M, et al. An
9 10	682	efficient quantitation method of next-generation sequencing libraries by using
11		
12 13	683	MiSeq sequencer. Anal Biochem. 2014; 466: 27-9.
14 15	684	
16 17	685	33. Koshiba S, Motoike I, Kojima K, Hasegawa T, Shirota M, Saito T, et al.
18 19	686	The structural origin of metabolic quantitative diversity. Sci Rep. 2016; 6:
20 21	687	31463.
22 23	688	
24 25	689	34. Nishiumi S, Kobayashi T, Ikeda A, Yoshie T, Kibi M, Izumi Y, et al. A novel
26 27	600	corum metabolomics based diagnostic approach for colorectal concer. BLoS
28	690	serum metabolomics-based diagnostic approach for colorectal cancer. PLoS
29 30	691	One. 2012; 7: e40459.
31 32	692	
33 34	693	35. Nishiumi S, Kobayashi T, Kawana S, Unno Y, Sakai T, Okamoto K, et al.
35 36	694	Investigations in the possibility of early detection of colorectal cancer by gas
37 38	695	chromatography/triple-quadrupole mass spectrometry. Oncotarget. 2017; 8,
39 40	696	17115-17126.
41 42 43	697	
44 45	698	36. Sato Y, Yamagishi J, Yamashita R, Shinozaki N, Ye B, Yamada T, et al.
46 47	699	Inter-Individual Differences in the Oral Bacteriome Are Greater than Intra-Day
48 49	700	Fluctuations in Individuals. PLoS One. 2015;10: e0131607.
50 51	701	
52 53	702	37. Watanabe K, Naruse K, Tanaka K, Metoki H, Suzuki Y. Outline of definition
54 55		
56	703	and classification of pregnancy induced hypertension (PIH). Hypertens Res
57 58		29
59 60		For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 31 of 35

1		
2 3 4	704	Pregnancy 2013; 1: 3–4.
5	705	
7 8 9 10 11 12	706	38. IADPSG Consensus Panel: International association of diabetes and
	707	pregnancy study groups recommendations on the diagnosis and classification
	708	of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682.
13 14	709	
15 16 17	710	39. #PRSice PRSice: Polygenic Risk Score software. Euesden J, Lewis CM,
18 19	711	O'Reilly PF, Bioinformatics, 2015; 31:1466-8.
20 21	712	
22 23	713	40. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for
24 25	714	the Risk of Preterm Labor: An Update. Clin Lab Med. 2016; 36: 369-83.
26 27 28	715	
29 30	716	41. Al-Rubaie Z, Askie LM, Ray JG, Hudson HM, Lord SJ. The performance of
31 32	717	risk prediction models for pre-eclampsia using routinely collected maternal
33 34	718	characteristics and comparison with models that include specialised tests and
35 36	719	with clinical guideline decision rules: a systematic review. BJOG. 2016;
37 38	720	123:1441-1452.
39 40	721	
41 42 43	722	42. Koullali B, Oudijk MA, Nijman TA, Mol BW, Pajkrt E. Risk assessment and
44 45	723	management to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21:
46 47	724	80-8.
48 49	725	
50 51	726	43. Henderson JT, Thompson JH, Burda BU, Cantor A. Preeclampsia
52 53	727	Screening: Evidence Report and Systematic Review for the US Preventive
54 55	728	Services Task Force. JAMA. 2017; 317: 1668-1683.
56 57 58		30
59 60		For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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.
	3

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.
32
For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml



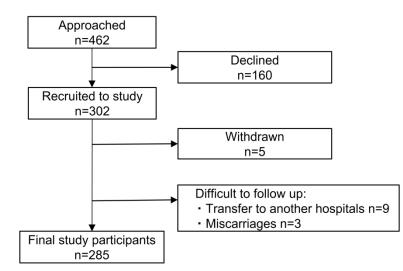
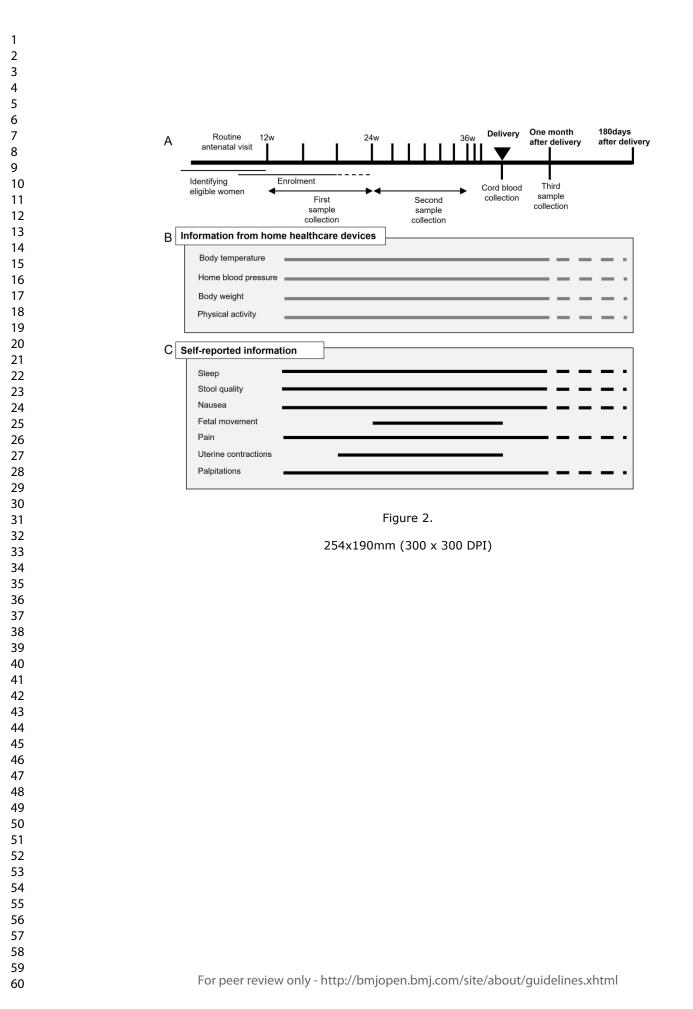
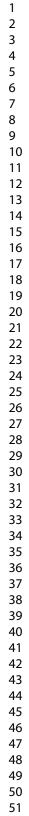


Figure 1.

254x190mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml





(%)

100

80

60

40

20

0

Physical

activity

Body weight

Body

temperature

- 56 57
- 58 59
- 60



Morning

pressure

home blood Home blood

Evening

pressure

Sleep quality Condition of

stool, pain,

nausea

uterine

contractions, palpitations

Fetal

movement

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; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; 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 Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University Shibata, Tomoko; Tohoku Medical Megabank Organization, Tohoku University Kawashima, Junko; Tohoku Medical Megabank Organization, Tohoku University Kasuoka, Fumiki; Tohoku Medical Megabank Organization, Tohoku University Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University Igarashi-Takai, Takako ; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical Megabank Organization, Tohoku University Fuse, Nobuo; 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

SCHOLARONE[™] Manuscripts

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Sendai, Miyagi, 980-8574, Japan.

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 Yamashita¹, Takafumi Yamauchi^{1,3}, Daisuke Saigusa¹, Maiko Wagata^{1,2}, Taku Obara¹, Mami Ishikuro¹, Yoshiki Tsunemoto³, Yuki Harada¹, Tomoko F.Shibata¹, Takahiro Mimori¹, Junko Kawashima¹, Fumiki Katsuoka¹, Takako Igarashi-Takai¹, Soichi Ogishima¹, Hirohito Metoki⁴, Hiroaki Hashizume¹, Nobuo Fuse^{1,2}, Naoko Minegishi¹, Seizo Koshiba¹, Osamu Tanabe^{1,5}, Shinichi Kuriyama^{1,2}, Kengo Kinoshita¹, Shigeo Kure^{1,2}, Nobuo Yaegashi^{1,6}, Masayuki Yamamoto^{1,2}, Satoshi Hiyama³, 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,

- 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-

$\begin{array}{c}1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\2\\3\\14\\15\\16\\17\\18\\9\\02\\1\\2\\23\\24\\25\\6\\7\\28\\9\\30\\3\\3\\4\\5\\36\\37\\38\\90\\41\\4\\2\\44\\45\\46\\7\\8\\2\\6\\7\\8\\9\\1\\2\\3\\3\\4\\5\\3\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\8\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\8\\8\\1\\2\\8\\1\\2\\8\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\1\\2$	
42 43 44 45 46 47	

25	8574, Japan.
26	
27	
28	Email addresses:
29	*Junichi Sugawara: jsugawara@med.tohoku.ac.jp
30	Daisuke Ochi: ochi@megabank.tohoku.ac.jp
31	Riu Yamashita: ryamasi@megabank.tohoku.ac.jp
32	Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp
33	Daisuke Saigusa: saigusa@m.tohoku.ac.jp
34	Maiko Wagata: wagata@med.tohoku.ac.jp
35	Taku Obara: obara-t@hosp.tohoku.ac.jp
36	Mami Ishikuro: m_ishikuro@med.tohoku.ac.jp
37	Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp
38	Yuki Harada: harada@megabank.tohoku.ac.jp
39	Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp
40	Takahiro Mimori: mimori@megabank.tohoku.ac.jp
41	Junko Kawashima: kawashima@dent.tohoku.ac.jp
42	Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp
43	Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp
44	Soichi Ogishima: ogishima@megabank.tohoku.ac.jp
45	Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp
46	Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp
47	Nobuo Fuse: fusen@megabank.tohoku.ac.jp
48	Naoko Minegishi: nmine@med.tohoku.ac.jp
49	Seizo Koshiba: koshiba@megabank.tohoku.ac.jp

1		
3 4 5	50	Osamu Tanabe: otanabe@rerf.or.jp
6 7	51	Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp
8 9	52	Kengo Kinoshita: kengo@ecei.tohoku.ac.jp
10 11	53	Shigeo Kure: kure@med.tohoku.ac.jp
12 13 14	54	Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp
15 16	55	Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp
17 18	56	Satoshi Hiyama: hiyamas@nttdocomo.com
19 20 21	57	*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp
21 22 23	58	
24 25	59	*Corresponding authors:
26 27 28 29 30 31 32 33 34 35 36 37 38 39	60	Junichi Sugawara: jsugawara@med.tohoku.ac.jp
	61	and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp
	62	Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi,
	63	Aoba-ku, 980-8573, Sendai, Japan. Phone: +81-22-273-6283
	64	
	65	Word count: 5314 words
40 41	66	Key words: lifelog, multi-omics analysis, prediction, complicated pregnancy
42 43 44	67	
45 46	68	
47 48	69	
49 50 51	70	
52 53	71	
54 55	72	
56 57 58	73	
59 60	74	
		0

Page 5 of 43

Abstract

BMJ Open

ed,	MJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.
sing	rom http://bmjopen.bmj.com/ on June 11, 2025 at Agen (ABES). and data mining, AI training, and similar technologies.
y of	mj.com raining
d	/ on Ju ا, and s
ody	ıne 11, similar
me	2025 at technol
ber	Agence Biblic ogies.
	ographique d
4	le I Enseignemer
	t

ΒM

76	Purpose: A prospective cohort study for pregnant women, the Maternity Log
77	study (MLOG), was designed to construct a time-course high-resolution
78	reference catalog of bioinformatic data in pregnancy and explore the
79	associations between genomic and environmental factors and the onset of
80	pregnancy complications, such as hypertensive disorders of pregnancy,
81	gestational diabetes mellitus, and preterm labor, using continuous lifestyle
82	monitoring combined with multi-omics data on the genome, transcriptome,
83	proteome, metabolome, and microbiome.
84	Participants: Pregnant women were recruited at the timing of first routine
85	antenatal visits at Tohoku University Hospital, Sendai, Japan between
86	September 2015 and November 2016. Of the eligible women who were invited,
87	65.4% agreed to participate, and a total of 302 women were enrolled. The
88	inclusion criteria were age \geq 20 years and the ability to access the internet using
89	a smartphone in the Japanese language.
90	Findings to date:
91	Study participants uploaded daily general health information including quality of
92	sleep, condition of bowel movements, and the presence of nausea, pain, and
93	uterine contractions. Participants also collected physiologic data, such as body

94 weight, blood pressure, heart rate, and body temperature, using multiple home

- 95 healthcare devices. The mean upload rate for each lifelog item was ranging
- 96 from 67.4 % (fetal movement) to 85.3% (physical activity) and the total number
- 97 of data points was over 6 million. Biospecimens, including maternal plasma,
- 98 serum, urine, saliva, dental plaque, and cord blood, were collected for multi-
- 99 omics analysis.

100 Future plans:

Lifelog and multi-omics data will be used to construct a time-course highresolution 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.

107 Strengths and limitations of this study:

- 108 This is the first study designed to collect longitudinal lifelog information
- 109 through healthcare devices, self-administered questionnaires using
- 110 smartphones, and varieties of biospecimens throughout pregnancy.
- 111 Longitudinal, continuous, individual lifelog data with a high acquisition rate
- 112 will enable us to assess dynamic physiological changes throughout
- 113 pregnancy. Multi-omics data will make it possible to understand the complex
- 114 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
- 117 high-risk populations. Therefore, the results might not be applicable to the
 - 118 general populations.
- 119 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
 - 122 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

1		
1 2 3 4		
3		
4	125	pregnancy complications.
5 6		
6 7	126	
8		
8 9	127	
10	100	
11	128	
12	400	
13 14	129	
15	120	
16	130	
17	101	
18	131	
19 20	132	
20 21	132	
22	133	
23	100	
24	134	
25	104	
26 27	135	
28		
29	136	
30		
31	137	
32		
33 34	138	
35		
36	139	
37		
38	140	
39 40		
41	141	
42		
43	142	
44 45		
45 46	143	
47	A A A	
48	144	
49	115	
50 51	145	
51 52	146	
53	140	
54	147	
55	171	
56 57	148	
57 58	ITU	
59	149	
60		
		6

150 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 multiomics analyses and objective data on environmental factors, including daily lifelog data, in pregnant women. This study may demonstrate correlations between specific lifelog patterns and pregnancy related physiological changes, such as blood pressure, gestational weight gain, and onset of obstetric diseases. Furthermore, studies on associations among lifelog patterns, plasma and urine metabolomes, transcriptomes, and genomic variations may reveal relationships among 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.

Page 9 of 43

1 2		
3 4	175	COHORT DESCRIPTION
5 6 7	176	Study setting
8 9	177	The aim of the MLOG study is to construct a time-course high-resolution
10 11	178	reference catalog of bioinformatic data in pregnancy and thereby develop
12 13 14	179	methods for early prediction of obstetric complications, through integrated
15 16 17 18	180	analysis of daily lifelogs and multi-omics data, <i>i.e.</i> , maternal genomes,
	181	transcriptomes, metabolomes, and oral microbiomes.
19 20 21	182	The MLOG study is a prospective, add-on cohort study, built on a birth- and 3-
21 22 23	183	generation cohort study established by the Tohoku Medical Megabank
24 25	184	Organization (TMM BirThree Cohort Study) [8] in order to elucidate the
26 27	185	mechanisms of complicated multifactorial diseases in mothers and children in
28 29 30 31 32 33 34 35 36 37 38 39 40 41	186	the wake of the Great East Japan Earthquake in 2011. Epidemiological data
	187	from extensive questionnaire surveys and accurate clinical records, including
	188	birth outcomes, can be abstracted from the integrated biobank of the Tohoku
	189	Medical Megabank Organization (ToMMo) [8]. TMM BirThree Cohort Study was
	190	started in July 2013 in one obstetric clinic and expanded throughout Miyagi
	191	Prefecture, and approximately 50 obstetric clinics and hospitals (including
42 43 44	192	Tohoku University Hospital) participated in the recruiting process. We planned
44 45 46	193	to recruit 20,000 pregnant women as probands, and her family members from
47 48	194	three generations, a total of over 70,000 participants [8]. Written informed
49 50	195	consent was obtained from all participants by the genome medical research
51 52 53	196	coordinators (GMRCs). The MLOG study was conducted in accordance with the
54 55	197	Declaration of Helsinki and approved by the ethics committees of Graduate
56 57	198	School of Medicine (2014-1-704) and ToMMo (22017-1-085), Tohoku University
58 59	199	under a collaborative research agreement among ToMMo, Tohoku University
60		8

BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

2 3		
4 5	200	and NTT DoCoMo, Inc. (Tokyo, Japan).
6 7	201	
8 9	202	Patient and public involvement
10 11	203	Patients or the public were not directly involved in the development of the
12 13 14	204	research question or the design of the study. The main results will be made
15 16	205	available in the public domain.
17 18	206	
19 20	207	Participants
21 22 23	208	Participants were recruited at a first routine antenatal visit at Tohoku University
24 25	209	Hospital, Sendai, Japan between September 2015 and November 2016. A
26 27	210	flowchart of the recruitment process is shown in Figure 1. GMRCs at Tohoku
28 29 30	211	University Hospital approached eligible pregnant women for TMM BirThree
31 32	212	Cohort Study (n= 631), and patients who already agreed to participate in TMM
33 34	213	BirThree Cohort Study (n=513) were assessed for eligibility for the MLOG study.
35 36	214	Finally, 462 pregnant women were asked to provide informed consent for the
37 38 39	215	MLOG study. A total of 302 women were enrolled. The inclusion criteria were
40 41	216	the age \geq 20 years and the ability to access the internet using a smartphone in
42 43	217	the Japanese language. Participants were excluded after enrollment if
44 45 46	218	termination of pregnancy, abortion, or transfer to another institution for
40 47 48	219	emergency care occurred before delivery, or if they withdrew consent for any
49 50	220	reason.
51 52	221	
53 54 55	222	Outline of study protocol
56 57	223	The study protocol consisted of blood and urine sampling, saliva and dental
58 59	224	plaque sampling, self-administered daily lifelog data collection, and data upload
60		9

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.

233 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.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

2		
3 4 5	250	Saliva and dental plaque sampling
5 6 7	251	Samples of saliva and dental plaque were collected 3 times from each
8 9	252	participant, at the same time points as blood collection. Approximately 3 mL of
10 11	253	saliva was collected using a 50-mL conical centrifuge tube (Corning, Inc.,
12 13 14	254	Corning, NY, USA) and stored at -80°C until analysis. Dental plaque was
15 16	255	sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM
17 18	256	EDTA; pH, 8.0), and immediately stored at -80°C until the time of sample
19 20	257	processing.
21 22 23	258	
24 25	259	Lifelog data collection
26 27	260	Based on previous publications on the utility for risk assessment of pregnancy-
28 29 30	261	related diseases, we selected several lifelog parameters to employ in this study,
31 32	262	<i>i.e.</i> , body temperature [10], home blood pressure [11], body weight [12], and
33 34	263	physical activity (calorie expenditure) [13], as well as self-administered
35 36 27	264	information such as sleep quality [14], condition of stool [15], severity of nausea
37 38 39	265	[16], fetal movement [17], severity of pain [18], uterine contractions [19], and
40 41	266	palpitations [20]. Body temperature, home blood pressure, body weight, and
42 43	267	physical activity were uploaded from multiple healthcare devices through a
44 45 46	268	smartphone. The self-administered information described above was input
47 48	269	manually on mobile applications created for this study.
49 50	270	Data collection was started after obtaining informed consent and after giving
51 52 53	271	detailed instructions for the use of the healthcare devices. These applications
53 54 55	272	tracked quality of sleep; condition of stool using the Bristol Scale [21-23];
56 57	273	severity of nausea using the Pregnancy-Unique Quantification of Emesis and
58 59	274	nausea (PUQE) score [24,25]; headache, toothache, lumbago, and upper and
60		11

Page 13 of 43

BMJ Open

1 2		
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	275	lower abdominal pain using a numerical rating scale (NRS) score; the number
	276	of perceived uterine contractions; palpitations; and fetal movement using a
	277	modified count-to-10 fetal movement chart [26,27].
	278	Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction
13	279	(ranked from satisfied to poor using a numeric scale of 0-4), and the number of
15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34	280	nocturnal awakenings (0-6).
	281	The Bristol stool form scale was originally developed to assess constipation
	282	and diarrhea [21, 22], and its use has been spread widely to evaluate functional
	283	bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types
	284	according to cohesion and surface cracking [21, 22].
	285	The PUQE score [24, 25] was developed to estimate the severity of nausea
	286	and vomiting in pregnancy and quantifies the number of daily vomiting and
	287	retching episodes and the length of nausea in hours (over the preceding 12 h).
	288	The total score ranges from 3 (no symptoms) to 15, and higher scores are
35 36 37	289	correlated with increasing severity of nausea and vomiting [24, 25].
38 39	290	In the NRS score for headache, toothache, lumbago, and upper and lower
40 41	291	abdominal pain, the total score ranges from 0 (no pain) to 10 (maximum ever
42 43 44	292	experienced).
44 45 46	293	Uterine contractions and palpitations were evaluated using definitions
47 48	294	determined for the current study. Uterine contractions were assessed using the
49 50	295	number of perceived contractions per day, ranging from 0 to more than 5. The
51 52 53	296	count-to-10 method was originally developed to assess fetal well-being by
54 55	297	recording the time, in minutes, required to count 10 fetal movements [26]. More
56 57	298	recently, a modified count-to-10 method has been proposed: pregnant women
58 59 60	299	are advised to start counting when they feel the first movement, then record the
00		12

 ABMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement

 Superieur (ABES) .

 Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

 Page

 Page

BMJ Open

> 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

BMJ Open

1 2		
3 4	325	Hospital. Epidemiological data, including extensive questionnaire surveys by
5 6 7	326	TMM BirThree Cohort Study can be obtained from the ToMMo integrated
7 8 9	327	biobank [8].
10 11	328	
12 13	329	Database
14 15 16	330	A customized laboratory information management system (LIMS) was
17 18	331	established to track all biospecimens. All data were transferred to the TMM
19 20	332	integrated database after 2-step anonymisation in a linkable fashion.
21 22 23	333	Data handling was strictly regulated under HIPAA (Health Insurance Portability
24 25	334	and Accountability Act of 1996, United States Security and Privacy Rules) [28,
26 27	335	29] and the Act on the Protection of Personal Information [30]. Security control
28 29	336	at our facility has been described previously [31].
30 31 32	337	
33 34	338	Omics analysis
35 36	339	Whole-genome sequencing
37 38 39	340	To minimize amplification bias, we adopted a PCR-free library preparation
40 41	341	method. After performing library quality control using the quantitative MiSeq
42 43	342	method [32], libraries were sequenced on HiSeq 2500 Sequencing System
44 45	343	(Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We
46 47 48	344	generated the sequencing data at over 12.5x coverage on average, and we
49 50	345	identified variants using the alignment tool BWA-MEM (ver. 0.7.5a-r405) with
51 52	346	the default option. Single nucleotide variants (SNVs) and indels were jointly
53 54 55	347	called across all samples using Genome Analysis Tool Kit's HaplotypeCaller
56 57	348	(ver. 3.8). Default filters were applied to SNV and indel calls using the GATK's
58 59	349	Variant Quality Score Recalibration (VQSR) approach. The human reference
60		14

BMJ Open

350	genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC_007605
351	(Human Gamma Herpesvirus 4). The complete fasta file named
352	hg19_tommo_v2.fa is available from iJGVD website
353	(http://ijgvd.megabank.tohoku.ac.jp) [33]. For the quality assurance, we have
354	checked the ratio of the bases with the phred quality score over 30, the total
355	variant numbers in each chromosome, and the ratio of transitions to
356	transversions for a pair of sequences.
357	
358	Transcriptome
359	Whole blood was collected using the PAXgene® RNA tube, which is widely
360	used for transcriptome analysis. After storage at -80°C, total RNA was purified
361	with PAXgene Blood RNA Kit $^{ m I\!R}$ (Qiagen, VenIo, The Netherlands) using
362	QiaSymphony $^{\ensuremath{\mathbb{R}}}$ (Qiagen). Total RNA was reverse-transcribed using an oligo-dT
363	primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.)
364	for library preparation for sequencing with HiSeq 2500 Sequencing System. For
365	the quality assurance, we randomly selected 11 samples in one batch (usually
366	48 samples) and checked an RNA integrity number (RIN) (or an RIN equivalent)
367	using BioAnalyzer $^{ m I\!R}$ or Tape Station $^{ m I\!R}$ (both from Agilent Technologies, Santa
368	Clara, CA, USA). The batch with RIN (or an RIN equivalent) higher than 7.0 for
369	all tested samples was used for the downstream analysis. The minimum
370	threshold for the total sequence reads for each sample was set to thirty millions.
371	For computing a series of quality control metrics for RNA-seq data, RNA-SeQC
372	was used to check the quality of sequence reads [34].
373	

374 Plasma and urine metabolome

1 2	
3 4 5	375
6 7	376
8 9	377
10 11	378
12 13 14	379
15 16	380
17 18	381
19 20 21	382
21 22 23	383
24 25	384
26 27	385
28 29 30	386
31 32	387
33 34	388
35 36 37	389
38 39	390
40 41	391
42 43 44	392
44 45 46	393
47 48	394
49 50	395
51 52 53	396
54 55	397
56 57	398
58 59 60	399
00	

375	Nuclear magnetic resonance (NMR) spectroscopy
376	All NMR measurements for metabolome analysis were conducted at 298 K on a
377	Bruker Avance 600 MHz spectrometer equipped with a SampleJet sample
378	changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional
379	nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-
380	Meiboom-Gill (CPMG) spectra were obtained for each plasma or urine sample.
381	All spectra for plasma or urine samples were acquired using 16 scans and 32 k
382	of complex data points. All data were analyzed using the TopSpin 3.5 (Bruker
383	Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta,
384	Canada) programs. All spectra were referenced to an internal standard (DSS-
385	d6). As necessary, those spectra were aligned using hierarchical cluster-based
386	peak alignment method, which is implemented as an R package called "speaq"
387	[36].
388	
389	Gas chromatography-tandem mass spectrometry (GC-MS/MS)
390	Sample preparation for plasma and urine (50 μ L each) was performed using a
391	Microlab STARlet robot system (Hamilton, Reno, NV, USA) followed by the
392	methods previously reported by Nishiumi [37, 38]. The resulting deproteinized
393	and derivatized supernatant (1 μL) was subjected to GC-MS/MS, performed on
394	a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound
395	separation was performed using a fused silica capillary column (BPX-5; 30 m $ imes$
396	0.25 mm inner diameter; film thickness, 0.25 μ m; 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

3	
4	
5	
6 7	
, 8	
9	
10	
11	
12 13	
14	
15	
16 17	
17	
18 19	
20	
21	
22	
23 24	
24 25	
26	
27	
28	
29 30	
31	
32	
33	
34 35	
36	
37	
38	
39 40	
40 41	
42	
43	
44	
45 46	
47	
48	
49	
50 51	
52	
53	
54	
55 56	
56 57	
58	
59	
60	

1 2

parameters, and retention index employed for the metabolite measurement. The 400 401 database used in this study included data on 475 peaks from 334 metabolites. 402 All peaks of metabolites detected from each sample was annotated and analyzed using Traverse MS® (Reifycs Inc., Tokyo, Japan). Then, two types of 403 404 normalization were performed to these annotated metabolites. The first 405 normalization was performed using the peak of 2-isopropylmalic acid as an internal standard which was added to each sample before analysis with GC-406 407 MS/MS. Then the second normalization was performed using quality control (QC) samples which were injected after every 12 study samples according to 408 409 the RQC normalization methods [39]. Normalized values of each metabolite in 410 the QC samples were assessed by calculating coefficients of variation (CVs), 411 and metabolites with CVs over 20% were eliminated.

412

413 Oral Microbiome

Analysis of oral microbiome was conducted by previously reported protocols 414 415 [40]. In brief, saliva was collected in a 50-mL tube. Dental plaque was sampled 416 by participants by brushing teeth with a sterilized toothbrush, and then 417 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 418 419 plaque by standard glass bead-based homogenization and subsequent 420 purification with a silica-membrane spin-column using PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). DNA was eluted from the spin 421 422 column with 30-µL RNase-free water (Takara Bio, Inc., Shiga, Japan), and 423 stored at -20°C after determining the amount and purity of DNA with a Nanodrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). 424

1

BMJ Open

ן ר	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20 21	
21	
//	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37 38	
39 40	
40 41	
41	
42 43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

425 Using DNA extracted from saliva or dental plaque as a template, a part of the 426 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 427 sequencing using MiSeq System with MiSeq Sequencing Reagent Kit, v3 428 429 (Illumina, Inc.) according to the manufacturer's instructions. For the quality 430 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 431 432 eliminate outliers.

433

434 Outcomes

The following obstetric complications represented the primary outcomes. 435 Gestational age was confirmed by measuring fetal crown rump length from 9 to 436 437 13 weeks of gestation using transvaginal ultrasound. HDP was defined as 438 gestational hypertension, preeclampsia, superimposed preeclampsia, or chronic hypertension [41,42]. Preterm birth was defined as spontaneous preterm labor, 439 440 medically induced preterm labor, or preterm premature rupture of membranes resulting in preterm birth at less than 37 weeks of gestation. GDM was 441 diagnosed according to the International Association of the Diabetes and 442 Pregnancy Study Groups (IADPSG) criteria [43]. The secondary outcomes were 443 444 maternal body weight, blood pressure, physical activity, lifestyle changes, 445 perinatal mental disorders, fetal growth, fetal movement, and birth weight. 446 Sample size calculation 447 448 At this time, there is little reliable evidence to demonstrate how time-dependent trends of longitudinal dense data would differ by pregnancy outcomes. 449

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.

461 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

Page 21 of 43

BMJ Open

475 analysis, the data might be collapsed with time scale, e.g. taking trimmed mean476 or median for each week, month, or trimester.

478 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.

9 499 Standard analyses would be also applicable for the selected variables and

BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

50	0	extracted factors (features). The association of outcomes with each feature will
50	1	be analyzed using statistical hypothesis tests such as Welch's t-test, Fisher's
50	2	exact test, the Chi-square test, and others as appropriate. Multiple logistic
50	3	regression modeling will be used to adjust for confounders and to assess
50	4	whether each feature or combination of features can be used to predict
50	5	outcomes. Stepwise selection algorithms or regularized algorithms (e.g.,
50	6	LASSO, ridge regression, or elastic net) will be used to select the optimal
50	7	number of contributing features that maximize the predictive power using the
50	8	leave-1-out cross validation or K-fold cross validation methods.
50	9	Individual genetic features may have an effect on outcomes; therefore, some
51	0	aggregated genetic risk score should be included in the prediction model. For
51	1	example, SNVs, including rare variants in or around a chromosome region of a
51	2	known or estimated risk gene, could be aggregated by considering their impacts
51	3	on biological function of the gene or their minor allele frequencies in the
51	4	population. However, this study is limited in the number of study participants,
51	5	and the aggregated risk score might therefore contribute only slightly to the
51	6	predictive power. To create a more reliable risk score, the estimates from other
51	7	large-scale cohort data using polygenic score tools, e.g., PRSice [55], could be
51	8	used for this study.
51	9	
52	0	FINDINGS TO DATE
52	1	Clinical background

⁴522 A total of 302 women were enrolled, and the mean gestational weeks of 7523 recruitment was 16.4 \pm 4.9 weeks (mean \pm SD). A total of 285 participants have 7524 been followed up to delivery; their baseline clinical characteristics are described

Page 23 of 43

BMJ Open

2		
3 4	525	in Table 1. The mean maternal age at delivery was 33.3 ± 4.9 years. As for
5 6 7	526	educational levels, 62% of the participants were high school graduates with or
8 9	527	without vocational college education, and 21% had a college degree. The
10 11	528	majority were employed (65%) in early pregnancy, and about 40% had a high
12 13 14	529	household income (over 6 million yen per year). Approximately 42% of the
15 16	530	participants were over 35 years of age, 51% were parous, and 22% were
17 18	531	overweight or obese by their prepregnancy body mass indices (BMI \ge 25 kg/m ²).
19 20 21	532	Overall, 8.4% of the participants had HDP, and 5.6% underwent spontaneous
21 22 23	533	preterm birth. On average, infants were delivered at 38.0 ± 2.3 weeks of
24 25	534	gestation with a mean birth weight of 2907 \pm 572 g. The rate of low birth weight
26 27	535	was 18%. Mean gestational weeks of the first and second blood sampling were
28 29 30	536	17.0 \pm 5.0 and 27.5 \pm 2.5, respectively. The third blood sampling was performed
31 32	537	at 31.1 \pm 3.0 days after delivery on average. The length of enrollment ranged
33 34	538	from 90 to 396 days with a mean of 216 \pm 61 days.
35 36	539	

Characteristics	Value
Maternal (n = 285)	
 Age at delivery, y, mean (SD) 	33.3 (± 4.9)
 Age at delivery, y, n (%) 20-24 25-29 30-34 35-39 40-44 45-49 	12 (4.2) 45 (15.8) 107 (37.5) 90 (31.6) 30 (10.5) 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)

	BMJ C	Dpen	Page 24 of 43
1			Page 24 of 43 BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 Protected by copyright, inc
2 3 4	College degree and above	17 (21.0)	n: first
5			publ
6 7	Others	1 (1.2)	ishe
8 9	Data not available	204	d as
10			10.1
11 12	Occupation (n=270) n (%)		136/k
13 14	Housewife or unemployed	93 (34.4)	Prote
15	Employed	175 (64.8)	cted
16 17			018- c
18	Student	2 (0.7)	0259 оруг
19 20			ight,
21 22	Annual household income, yen (n=248) n (%)		n 9 F
23	< 2 million	17 (6.9)	bmjopen-2018-025939 on 9 February 2019. Downloaded from t Superieur (ABE Protected by copyright, including for uses related to text and
24 25	2-4 million	59 (23.8)	ary 2 y for
26 27	4-6 million		2019. uses
28		73 (29.4)	Dow اعاد
29 30	6-8 million	51 (20.6)	wnloaded f Superieur ated to text
31	8-10 million	22 (8.9)	o tex
32 33	> 10 million	26 (10.5)	from (ABE t and
34 35			
36	・Parity, n (%)		http://bmjopen.bmj.com/ on June 11, 2025 at Agen S) . data mining, Al training, and similar technologies
37 38	0	140 (49.1)	ning,
39	1 ≥ 2	93 (32.6) 52 (18.2)	Al t
40 41		22.7 (1.5.4)	aini.co
42 43	 Prepregnancy BMI*, kg/m², mean (SD) Prepregnancy BMI, kg/m², n (%) 	22.7 (± 5.1)	າ <u>ຫ</u> , a
44	< 18.5	36 (12.6)	n n u si
45 46	18.5-24.9 25.0-29.9	186 (65.3) 34 (11.9)	imila 1
47	≥ 30.0	29 (10.2)	r tec
48 49	· Contational weaks at delivery mean (SD)	38 0 (+ 3 3)	hnol
50 51	Gestational weeks at delivery, mean (SD)	38.0 (± 2.3)	: Age ogie
52	Mode of delivery, n (%)		s.
53 54	Noncesarean Cesarean	179 (62.8) 106 (37.2)	Bibli
55	Cesarean	100 (37.2)	iogra
56 57	Pregnancy complication, n (%)		ıphic
58 59	Hypertensive disorder of pregnancy Spontaneous preterm birth	24 (8.4) 16 (5.6)	lue d
60			е — Е
			23 se i
	-		 (1) - (1) - (2) -
	For peer review only - http://bmjopen.	omj.com/site/about/guidelines.xhtm	nent

2			
3			
4		Neonatal (n = 300)	
5 6		 Birth weight, g, mean (SD) 	2907 (± 572)
7		• Sex, n (%)	
8		Male	168 (56)
9		Female	132 (44)
10 11	E 1 1	Low-birth weight (< 2500 g), n (%)	54 (18)
12	541	*BMI, body mass index	
13	542		
14			
15 16	543	Data acquisition	
17			
18	544	The percentage of data uploads as of	June 2017 was calculated for the 285
19 20		final study a stining stor. East a she life to a	, its set the surplus of sets for so the
20 21	545	final study participants. For each lifelog	g item, the upload rate for each
22	546	participant was calculated from the tota	al number of days of actual unloads
23	540	participant was calculated from the tota	a number of days of actual uploads
24 25	547	divided by the number of days from en	rollment to delivery. The mean upload
25 26	-		
27	548	rate for each lifelog item was 85.3% (p	hysical activity), 82.1% (body weight),
28			
29 30	549	80.4% (body temperature), 78.0% (mo	rning home blood pressure), 71.6%
30 31		<i>.</i>	
32	550	(evening home blood pressure), 83.5%	(sleep quality), 82.1% (condition of stool,
33	EE 1	acvarity of pain, acvarity of payson, ut	ring contractions, and polnitations) and
34 35	551	seventy of pain, seventy of hausea, ute	erine contractions, and palpitations), and
36	552	67.4% (fetal movement) (Figure 3).	
37			
38	553		
39 40			
41	554	Number of data points	
42			
43 44	555	The total number of collected data poir	nts as of June 2017 was calculated for the
44 45	FFC	295 final study participants. The approx	vimate number of registered data points
46	556	200 mai study participants. The appro-	ximate number of registered data points
47	557	was 86 000 for body weight, 324 000 p	points for home diastolic and systolic
48 49	001		
49 50	558	blood pressure, 86 000 for physical act	tivity, and 74 000 for body temperature.
51			
52	559	When physical conditions such as stoo	ol condition, severity of pain, and fetal
53 54			
55	560	movement were combined, the total nu	Imber of data points was over 6 million.
56			
57	561		
58 59	562	STRENGTHS AND LIMITATIONS	
60	502		
			24

1 2	
3	
4	
5	
6	
7 8	
o 9	
10	
11	
12	
13	
14 15	
16	
17	
18	
19 20	
21	
22	
23	
24	
25	
26 27	
27	
29	
30	
31	
32 33	
33 34	
35	
36	
37	
38	
39 40	
41	
42	
43	
44 45	
45 46	
47	
48	
49	
50 51	
51 52	
53	
54	
55	
56	
57 58	
58 59	
60	

5	63	Herein, we have described the rationale, design, objective, data collection
5	64	methods, and interim results of the MLOG study. The study was launched in
5	65	September 2016, and baseline data collection ended in June 2017. A total of
5	66	285 participants uploaded lifelog data throughout pregnancy with a high data
5	67	acquisition rate and over 6 million total data points. Biospecimens for multi-
5	68	omics analysis were satisfactorily collected and all tracked by LIMS.
5	69	There are three noteworthy features in the MLOG study. First, it is a
5	70	prospective add-on cohort study based on TMM BirThree Cohort Study, with a
5	71	full series of epidemiological data and a highly structured follow-up system for
5	72	mothers, newborns, and families [8]. Second, we have successfully collected
5	73	longitudinal, continuous, individual lifelog data with a high acquisition rate,
5	74	which will enable us to assess dynamic changes in physiologic conditions
5	75	throughout pregnancy. Third, multi-omics data will make it possible to fully
5	76	understand the complex mechanisms of multifactorial pregnancy-related
5	77	diseases and to overcome the unpredictability of these complications.
5	78	Prediction models using clinical and epidemiological information and
5	79	circulating factors for pregnancy-related diseases have been developed
5	80	extensively [56], and risk-assessment approaches using clinical information
5	81	have also been developed [57, 58]. However, there is a lack of evidence for the
5	82	benefits of these predictive models for routine clinical use [59]. Once the
5	83	likelihood of a pregnancy-related disorder is estimated with high sensitivity and
5	84	specificity, evidence-based clinical interventions could reduce the rate of
5	85	maternal and neonatal morbidity and mortality [60]. Therefore, an early-
5	86	prediction algorithm that can be used with a high level of confidence is needed
5	87	to obtain better outcomes for patients with pregnancy complications.

Page 27 of 43

BMJ Open

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

⁴³BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies. Page PA

BMJ Open

2		
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	613	reference catalog of wellness and multi-omics data from pregnant women and
	614	thereby develop a personalized predictive model for pregnancy complications.
	615	Progressive data sharing and collaborative studies would make it possible to
	616	establish a standardized early-prediction method through large clinical trials.
	617	
	618	Author affiliations
	619	1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-
	620	machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.
	621	2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku,
24 25	622	Sendai, Miyagi, 980-8574, Japan.
26 27	623	3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,
28 29 30	624	Kanagawa, Japan 239-8536.
31 32	625	4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-
33 34	626	ku, Sendai, Miyagi 981-8558, Japan.
35 36 37 38 39 40 41	627	5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku,
	628	Hiroshima 732-0815, Japan.
	629	6 Tohoku University Hospital, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-
42 43	630	8574, Japan.
44 45 46	631	
47 48	632	
49 50	633	Acknowledgements
51 52 53 54 55 56 57 58 59 60	634	The authors would like to thank all the MLOG study participants, the staff of the
	635	Tohoku Medical Megabank Organization, Tohoku University (a full list of
	636	members is available at: http://www.megabank.tohoku.ac.jp/english/a161201/),
	637	and the Department of Obstetrics and Gynecology, Tohoku University Hospital,

Page 29 of 43

BMJ Open

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).

1 2		
3 4	663	
5 6 7 8 9 10 11	664	Competing interests
	665	This study was funded by NTT DoCoMo, Inc.
	666	Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT
12 13	667	DoCoMo, Inc. All other authors declare that they have no competing interests.
14 15 16 17	668	
18 19 20	669	Ethics approval and consent to participate
21 22 23 24 25 26 27 28 29 30 31	670	TMM BirThree Cohort Study was approved by the ethics committees of the
	671	Tohoku University (authorization numbers, 2013-4-103 and 2017-4-010). The
	672	MLOG study was approved by the ethics committees of the Graduate School of
	673	Medicine (2014-1-704) and the Tohoku Medical Megabank Organization (2017-
	674	1-085), Tohoku University. Written informed consent was obtained from all
32 33	675	participants.
34 35 36	676	
37 38	677	Provenance and peer review
39 40	678	Not commissioned; externally peer reviewed.
41 42 43	679	
44 45	680	Data sharing statement
46 47 48 49	681	The datasets used during the current study are available from the
	682	corresponding authors on reasonable request.
50 51 52	683	
53 54 55 56	684	
	685	
57 58 59 60	686	
		20

1 2		
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25	687	
	688	
	689	
	690	REFERENCES
	691	1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public
	692	health perspective. Diabetes Care. 2007;30 Suppl 2: S141-6. doi:
	693	10.2337/dc07-s206.
	694	
	695	2. Beck S, Wojdyla D, Say L, et al. The worldwide incidence of preterm birth: a
	696	systematic review of maternal mortality and morbidity. Bull World Health Organ.
26 27 28	697	2010; 88:31-8. doi: 10.2471/BLT.08.062554.
29 30	698	
31 32 33 34 35	699	3. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol.
	700	2009; 33:130-7. doi: 10.1053/j.semperi.2009.02.010.
36 37	701	
38 39	702	4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United States,
40 41 42	703	1980-2010: age-period-cohort analysis. BMJ. 2013;347: f 6564. doi:
42 43 44	704	10.1136/bmj.f6564.
45 46	705	
47 48	706	5. Waken RJ, de Las Fuentes L, Rao DC. A Review of the Genetics of
49 50 51 52 53 54 55 56 57 58 59 60	707	Hypertension with a Focus on Gene-Environment Interactions. Curr Hypertens
	708	Rep. 2017; 19:23. doi: 10.1007/s11906-017-0718-1.
	709	
	710	6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia /
	711	eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:

BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

1 2		
3 4 5	712	Elsevier; 2990: 51-72.
6 7	713	
8 9	714	7. Li X, Dunn J, Salins D, Zhou G, et al. Digital Health: Tracking Physiomes and
10 11	715	Activity Using Wearable Biosensors Reveals Useful Health Related Information.
12 13 14	716	PLoS Biol. 2017; 15: e2001402. doi: 10.1371/journal.pbio.2001402.
15 16	717	
17 18	718	8. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank
19 20 21	719	Project: Design and Mission. J Epidemiol. 2016; 26:493-511. doi:
21 22 23	720	10.2188/jea.JE20150268.
24 25	721	
26 27 28	722	9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical
28 29 30	723	Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo, Japan,
31 32	724	pp. 1–4, 2017 [in Japanese].
33 34	725	
35 36 37	726	10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the
38 39	727	thermoneutral zone: a longitudinal study of normal human pregnancy. Acta
40 41	728	Physiol (Oxf). 2011; 201: 467-74. doi: 10.1111/j.1748-1716.2010.02228.x.
42 43 44	729	
45 46	730	11. Metoki H, Ohkubo T, Watanabe Y, et al. Seasonal trends of blood pressure
47 48	731	during pregnancy in Japan: the babies and their parents' longitudinal
49 50	732	observation in Suzuki Memorial Hospital in Intrauterine Period study. J
51 52 53	733	Hypertens. 2008; 26: 2406-13. doi: 10.1097/HJH.0b013e32831364a7.
54 55	734	
56 57	735	12. Haugen M, Brantsæter AL, Winkvist A, et al. Associations of pre-pregnancy
58 59 60	736	body mass index and gestational weight gain with pregnancy outcome and
00		31

2		
3 4	737	postpartum weight retention: a prospective observational cohort study. BMC
5 6 7	738	Pregnancy Childbirth. 2014 Jun 11; 14: 201. doi: 10.1186/1471-2393-14-201.
, 8 9	739	
10 11	740	13. Sorensen TK, Williams MA, Lee IM, et al. Recreational physical activity
12 13	741	during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-
14 15 16	742	80. doi: 10.1161/01.HYP.0000072270.82815.91
17 18	743	
19 20	744	14. Reutrakul S, Zaidi N, Wroblewski K, et al. Sleep disturbances and their
21 22	745	relationship to glucose tolerance in pregnancy. Diabetes Care. 2011; 34: 2454-7.
23 24 25	746	doi: 10.2337/dc11-0780.
26 27	747	
28 29	748	15. Cornish J, Tan E, Teare J, et al. A meta-analysis on the influence of
30 31 32	749	inflammatory bowel disease on pregnancy. Gut. 2007; 56: 830-7. doi:
33 34	750	10.1136/gut.2006.108324.
35 36	751	
37 38	752	16. Huxley RR. Nausea and vomiting in early pregnancy: its role in placental
39 40	753	development. Obstet Gynecol. 2000; 95:779-82.
41 42 43	754	
44 45	755	17. Holm Tveit JV, Saastad E, Stray-Pedersen B, et al. Maternal characteristics
46 47	756	and pregnancy outcomes in women presenting with decreased fetal movements
48 49 50	757	in late pregnancy. Acta Obstet Gynecol Scand. 2009; 88: 1345-51. doi:
50 51 52	758	10.3109/00016340903348375.
53 54		10.3109/00010340903348373.
55 56	759	
57 58	760	18. Facchinetti F, Allais G, D'Amico R, et al. The relationship between
59 60	761	headache and preeclampsia: a case-control study. Eur J Obstet Gynecol
		32

2 3		
3 4 5	762	Reprod Biol. 2005; 121: 143-8. doi: 10.1016/j.ejogrb.2004.12.020.
6 7	763	
8 9	764	19. lams JD, Newman RB, Thom EA, et al; National Institute of Child Health
10 11	765	and Human Development Network of Maternal-Fetal Medicine Units. Frequency
12 13	766	of uterine contractions and the risk of spontaneous preterm delivery. N Engl J
14 15 16	767	Med. 2002; 346: 250-5. doi: 10.1056/NEJMoa002868
17 18	768	
19 20	769	20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular
21 22	770	system. Int J Cardiol. 2005; 98: 179-89. doi: 10.1016/j.ijcard.2003.10.028.
23 24 25	771	
26 27	772	21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit
28 29	773	time. Scand J Gastroenterol. 1997; 32: 920-4.
30 31	774	doi: 10.3109/00365529709011203
32 33 34	775	
35 36	776	22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical
37 38	777	practice and clinical research. Tech Coloproctol 2001; 5: 163–4. doi:
39 40	778	10.1007/s101510100019
41 42 43	779	
44 45	780	23. Longstreth GF, Thompson WG, Chey WD, et al. Functional Bowel Disorders.
46 47	781	Gastroenterology 2006; 130: 1480–91. doi: 10.1053/j.gastro.2005.11.061
48 49		Gastroenterology 2000, 130. 1480–91. doi: 10.1033/j.gastro.2003.11.001
50 51	782	24. Karan C. Baskavia D. Hard M. et al. Mathemiak DHOE (magnetic unique
52 53 54	783	24. Koren G, Boskovic R, Hard M, et al. Motherisk-PUQE (pregnancy-unique
55 56	784	quantification of emesis and nausea) scoring system for nausea and vomiting of
57 58	785	pregnancy. Am J Obstet Gynecol. 2002;186: S228-31.
59 60	786	
		33

2		
3 4 5	787	25. Koren G, Piwko C, Ahn E, et al. Validation studies of the Pregnancy Unique-
5 6 7	788	Quantification of Emesis (PUQE) scores. J Obstet Gynaecol. 2005; 25: 241-4.
8 9	789	doi: 10.1080/01443610500060651
10 11	790	
12 13 14	791	26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. Br
14 15 16	792	Med J 1976; 1:1305–7.
17 18	793	
19 20	794	27. Winje BA, Saastad E, Gunnes N, et al. Analysis of 'count-to-ten' fetal
21 22 23	795	movement charts: a prospective cohort study. BJOG. 2011;118: 1229-38. doi:
24 25	796	10.1111/j.1471-0528.2011.02993.x
26 27	797	
28 29 30	798	28. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach
31 32	799	Notification rules under the Health Information Technology for Economic and
33 34	800	Clinical Health Act and the Genetic Information Nondiscrimination Act; other
35 36	801	modifications to the HIPAA rules. Fed Regist. 2013; 78: 5565-702.
37 38 39	802	
40 41	803	29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-
42 43	804	191. US Statut Large. 1996; 110:1936-2103.
44 45 46	805	
47 48	806	30. Amended Act on the Protection of Personal Information.
49 50	807	https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.
51 52	808	pdf.
53 54 55	809	
56 57	810	31. Takai-Igarashi T, Kinoshita K, Nagasaki M, et al. Security controls in an
58 59	811	integrated Biobank to protect privacy in data sharing: rationale and study design.
60		34

2 3		
4 5	812	BMC Med Inform Decis Mak. 2017; 17:100. doi: 10.1186/s12911-017-0494-5
6 7	813	
8 9	814	32. Katsuoka F, Yokozawa J, Tsuda K, et al. An efficient quantitation method of
10 11 12	815	next-generation sequencing libraries by using MiSeq sequencer. Anal Biochem.
12 13 14	816	2014; 466: 27-9. doi: 10.1016/j.ab.2014.08.015
15 16	817	
17 18 19	818	33.Yamaguchi-Kabata Y, et al, Nariai N, Kawai Y, et al. iJGVD: an
20 21	819	integrative Japanese genome variation database based on whole-genome
22 23	820	sequencing. Hum Genome Var. 2015; 2:15050. doi: 10.1038/hgv.2015.50.
24 25	821	https://www.ncbi.nlm.nih.gov/pubmed/27081555
26 27 28	822	
29 30	823	34. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics
31 32	824	for quality control and process optimization. Bioinformatics. 2012; 28: 1530-
33 34 35	825	1532. doi: 10.1093/bioinformatics/bts196
36 37	826	
38 39	827	35. Koshiba S, Motoike I, Kojima K, et al. The structural origin of metabolic
40 41	828	quantitative diversity. Sci Rep. 2016; 6: 31463. doi: 10.1038/srep31463
42 43 44	829	
45 46	830	36. Vu TN, Valkenborg D, Smets K, et al. An integrated workflow for robust
47 48	831	alignment and simplified quantitative analysis of NMR spectrometry data. BMC
49 50 51	832	Bioinformatics. 2011; 12: 405. doi: 10.1186/1471-2105-12-405.
52 53	833	
54 55	834	37. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based
56 57	835	diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459. doi:
58 59 60	836	10.1371/journal.pone.0040459.
00		35

2			
3 4	837		
5 6 7	838	38. Nishiumi S, Kobayashi T, Kawana S, et al. Investigations in the possibility o	f
8 9	839	early detection of colorectal cancer by gas chromatography/triple-quadrupole	
10 11	840	mass spectrometry. Oncotarget. 2017; 8, 17115-17126. doi:	
12 13 14	841	10.18632/oncotarget.15081.	
15 16	842		
17 18	843	39. Saigusa D, Okamura Y, Motoike IN, et al. Establishment of Protocols for	
19 20 21	844	Global Metabolomics by LC-MS for Biomarker Discovery. PLoS One.	
22 23	845	2016;11(8): e0160555. doi: 10.1371/journal.pone.0160555.	
24 25	846		
26 27 28	847	40. Sato Y, Yamagishi J, Yamashita R, et al. Inter-Individual Differences in the	
28 29 30	848	Oral Bacteriome Are Greater than Intra-Day Fluctuations in Individuals. PLoS	
31 32	849	One. 2015;10: e0131607. doi: 10.1371/journal.pone.0131607.	
33 34 35	850		
35 36 37	851	41. Brown MA, Magee LA, Kenny LC, et al. Hypertensive Disorders of	
38 39	852	Pregnancy: ISSHP Classification, Diagnosis, and Management	
40 41	853	Recommendations for International Practice. Hypertension. 2018; 72: 24–43.	
42 43 44	854	doi: 10.1161/HYPERTENSIONAHA.117.10803.	
45 46	855		
47 48	856	42. Watanabe K, Naruse K, Tanaka K, et al. Outline of definition and	
49 50 51	857	classification of pregnancy induced hypertension (PIH). Hypertens Res	
52 53	858	Pregnancy 2013; 1: 3–4.	
54 55	859		
56 57 58	860	43. IADPSG Consensus Panel: International association of diabetes and	
58 59 60	861	pregnancy study groups recommendations on the diagnosis and classification	
		3	6

1 2		
3 4 5	862	of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682. doi:
5 6 7	863	10.2337/dc09-1848.
8 9	864	
10 11	865	44. Box GEP, Jenkins GM, Reinsel GC. Time series Analysis: Forecasting and
12 13 14	866	Control. 5th ed. New Jersey: Wiley; 2015.
15 16	867	
17 18	868	45. Brandt PT, Williams JT. Multiple Time Series Models, Thousand Oaks, CA:
19 20 21	869	Sage Publications, 2007.
21 22 23	870	
24 25	871	46. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data,
26 27	872	Journal of Statistical Software, 45 (7) 2011.
28 29 30	873	
31 32	874	47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by
33 34	875	surrogate variable analysis. PLoS Genet. 2007; 3: 1724-35. doi:
35 36 37	876	10.1371/journal.pgen.0030161
37 38 39	877	
40 41	878	48. Stegle O, Parts L, Piipari M, et al. Using probabilistic estimation of
42 43	879	expression residuals (PEER) to obtain increased power and interpretability of
44 45 46	880	gene expression analyses. Nat Protoc. 2012; 7: 500-7. doi:
47 48	881	10.1038/nprot.2011.457.
49 50	882	
51 52 53	883	49. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene
55 54 55	884	expression across human tissues. Nature. 2017; 550: 204-213. doi:
56 57	885	10.1038/nature24277.
58 59	886	
60		22

2		
3 4 5	887	50. Polgreen PM, Yang M, Kuntz JL, et al. Using oral vancomycin prescriptions
6 7	888	as a proxy measure for Clostridium difficile infections: a spatial and time series
8 9	889	analysis. Infect Control Hosp Epidemiol. 2011; 32: 723-6. doi: 10.1086/660858.
10 11	890	
12 13 14	891	51. McDowell IC, Manandhar D, Vockley CM, et al. Clustering gene expression
15 16	892	time series data using an infinite Gaussian process mixture model. PLoS
17 18	893	Comput Biol. 2018; 14: e1005896. doi: 10.1371/journal.pcbi.1005896.
19 20 21	894	
21 22 23	895	52. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of
24 25	896	Structured Time-Series. IEEE Trans Pattern Anal Mach Intell. 2015; 37: 383-93.
26 27	897	doi: 10.1109/TPAMI.2014.2318711.
28 29 30	898	
31 32	899	53. Rohart F, Gautier B, Singh A et al. mixOmics: An R package for 'omics
33 34	900	feature selection and multiple data integration. PLoS Comput Biol. 2017;13:
35 36 37	901	e1005752. doi: 10.1371/journal.pcbi.1005752.
37 38 39	902	
40 41	903	54. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a
42 43	904	framework for unsupervised integration of multi-omics data sets. Mol Syst Biol.
44 45 46	905	2018; 14: e8124. doi: 10.15252/msb.20178124.
47 48	906	
49 50	907	55. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software.
51 52 53	908	Bioinformatics, 2015; 31:1466-8. doi: 10.1093/bioinformatics/btu848.
53 54 55	909	
56 57	910	56. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for
58 59	911	the Risk of Preterm Labor: An Update. Clin Lab Med. 2016; 36: 369-83. doi:
60		20

ABMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement
 Superieur (ABES) .
 Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.
 Page

BMJ Open

57. Al-Rubaie Z, Askie LM, Ray JG, et al. The performance of risk prediction

models for pre-eclampsia using routinely collected maternal characteristics and

comparison with models that include specialised tests and with clinical guideline

58. Koullali B, Oudijk MA, Nijman TA, et al. Risk assessment and management

to prevent preterm birth. Semin Fetal Neonatal Med. 2016 ;21: 80-8. doi:

decision rules: a systematic review. BJOG. 2016; 123:1441-1452. doi:

2	
3 ⊿	
4 5	
6	
7	
8	
4 5 7 8 9 10	
10	
11 12	
13	
14	
15	
16	
17	
12 13 14 15 16 17 18 19	
19 20	
21	
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35	
23	
24	
25	
20 27	
28	
29	
30	
31	
32	
33 34	
35	
36 37	
37	
38	
39	
40 41	
41	
43	
44	
45	
46	
47 48	
40 49	
50	
51	
52	
53	
54 55	
55 56	
57	
58	
59	
60	

10.1016/j.cll.2016.01.019.

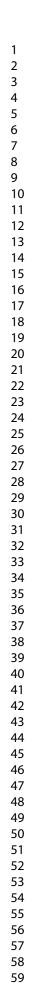
10.1111/1471-0528.

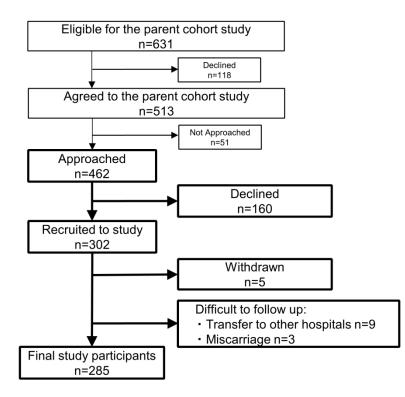
10.1016/j.siny.2016.01.005.

doi: 10.1038/nbt.3870.

1 2			
3 4	937		
5 6 7	938		
8 9	939		
10 11	940	FIGURE TITLES AND LEGENDS	
12 13 14	941		
15 16	942	Figure 1. Flowchart of Maternity Log Study (MLOG) participants	
17 18	943		
19 20 21	944	Figure 2. Overview of the MLOG study protocol	
21 22 23	945	A: Participant timeline for the MLOG study.	
24 25	946	B: Physiologic information collected using healthcare devices. Specific	
26 27	947	measures were uploaded each day from the time of enrollment (solid horizonta	al
28 29 30 31 32	948	lines). Participants had the option to continue uploading data until 180 days	
	949	after delivery (dashed horizontal lines).	
33 34	950	C: Daily lifelogs of self-reported information using a smartphone application.	
35 36 37	951	Basic lifelog information was input manually from the time of enrollment (solid	
38 39	952	horizontal lines). Participants had the option to continue uploading data until	
40 41	953	180 days after delivery (dashed horizontal lines). Fetal movement and uterine	
42 43 44	954	contractions were recorded from 24 and 20 weeks of gestation, respectively.	
44 45 46	955		
47 48	956	Figure 3. Data acquisition rate	
49 50	957	The mean data upload rate of specific measures was calculated from the total	
51 52 53	958	number of days of actual uploads divided by the number of days from	
54 55	959	enrollment to delivery for each participant.	
56 57			
58 59 60			
			40

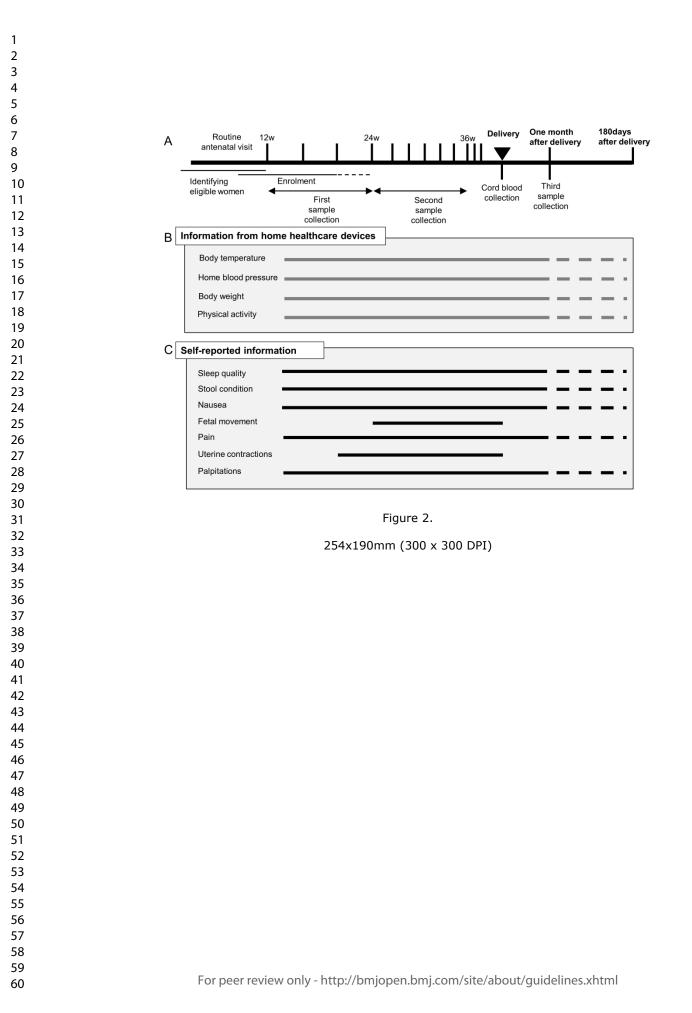
BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.





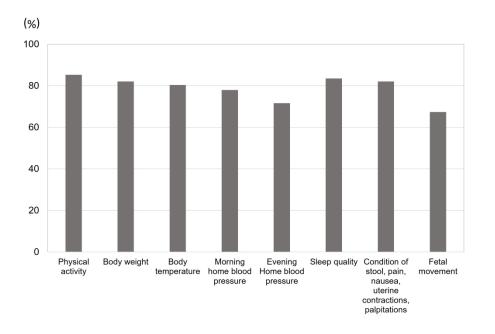


254x190mm (300 x 300 DPI)



pert first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . ge Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies. Pa

BMJ Open





254x190mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Research Laboratories, NTT DOCOMO, INC. Saigusa, Daisuke; Tohoku Medical Megabank Organization, Tohoku University; Tohoku Medical Megabank Organization, Tohoku University Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University Ishikuro, Mami; Tohoku Medical Megabank Organization, Tohoku University Ishikuro, 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 Metoki, Hirohito; Tohoku Medical Megabank Organization, Tohoku University Metoki, Hirohito; Tohoku Medical Megabank Organization, Tohoku University Fuse, Nobuo; Tohoku Medical Megabank Organization, Tohoku University Koshiba, Seizo; Tohoku Medical Megabank Organization, Tohoku University

pen: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

	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

SCHOLARONE[™] Manuscripts

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 Yamashita¹, Takafumi Yamauchi^{1,3}, Daisuke Saigusa¹, Maiko Wagata^{1,2}, Taku Obara¹, Mami Ishikuro¹, Yoshiki Tsunemoto³, Yuki Harada¹, Tomoko F.Shibata¹, Takahiro Mimori¹, Junko Kawashima¹, Fumiki Katsuoka¹, Takako Igarashi-Takai¹, Soichi Ogishima¹, Hirohito Metoki⁴, Hiroaki Hashizume¹, Nobuo Fuse^{1,2}, Naoko Minegishi¹, Seizo Koshiba¹, Osamu Tanabe^{1,5}, Shinichi Kuriyama^{1,2}, Kengo Kinoshita¹, Shigeo Kure^{1,2}, Nobuo Yaegashi^{1,6}, Masayuki Yamamoto^{1,2}, Satoshi Hiyama³, 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, Aobaku, 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-

$\begin{array}{c}1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\2\\3\\14\\15\\16\\17\\18\\9\\02\\1\\2\\23\\24\\25\\6\\7\\28\\9\\30\\3\\3\\4\\5\\36\\37\\38\\90\\41\\4\\2\\44\\45\\46\\7\\8\\2\\6\\7\\8\\9\\1\\2\\3\\3\\4\\5\\3\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\9\\0\\1\\2\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\9\\0\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\8\\1\\2\\3\\3\\4\\5\\6\\7\\8\\8\\8\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\8\\1\\2\\3\\1\\2\\8\\1\\2\\2\\1\\2\\2\\2\\2\\2\\2\\2\\2\\2\\2\\2\\2\\2$	
42 43 44 45 46 47	

25	8574, Japan.
26	
27	
28	Email addresses:
29	*Junichi Sugawara: jsugawara@med.tohoku.ac.jp
30	Daisuke Ochi: ochi@megabank.tohoku.ac.jp
31	Riu Yamashita: ryamasi@megabank.tohoku.ac.jp
32	Takafumi Yamauchi: t.yamauchi@megabank.tohoku.ac.jp
33	Daisuke Saigusa: saigusa@m.tohoku.ac.jp
34	Maiko Wagata: wagata@med.tohoku.ac.jp
35	Taku Obara: obara-t@hosp.tohoku.ac.jp
36	Mami Ishikuro: m_ishikuro@med.tohoku.ac.jp
37	Yoshiki Tsunemoto: yoshiki.tsunemoto@megabank.tohoku.ac.jp
38	Yuki Harada: harada@megabank.tohoku.ac.jp
39	Tomoko F. Shibata: tshibata@megabank.tohoku.ac.jp
40	Takahiro Mimori: mimori@megabank.tohoku.ac.jp
41	Junko Kawashima: kawashima@dent.tohoku.ac.jp
42	Fumiki Katsuoka: kfumiki@med.tohoku.ac.jp
43	Takako Igarashi-Takai: takai@megabank.tohoku.ac.jp
44	Soichi Ogishima: ogishima@megabank.tohoku.ac.jp
45	Hirohito Metoki: hmetoki@tohoku-mpu.ac.jp
46	Hiroaki Hashizume: hashizume@megabank.tohoku.ac.jp
47	Nobuo Fuse: fusen@megabank.tohoku.ac.jp
48	Naoko Minegishi: nmine@med.tohoku.ac.jp
49	Seizo Koshiba: koshiba@megabank.tohoku.ac.jp

1 2		
3 4	50	Osamu Tanabe: otanabe@rerf.or.jp
5 6 7	51	Shinichi Kuriyama: kuriyama@med.tohoku.ac.jp
8 9	52	Kengo Kinoshita: kengo@ecei.tohoku.ac.jp
10 11	53	Shigeo Kure: kure@med.tohoku.ac.jp
12 13 14	54	Nobuo Yaegashi: yaegashi@med.tohoku.ac.jp
15 16	55	Masayuki Yamamoto: masiyamamoto@med.tohoku.ac.jp
17 18	56	Satoshi Hiyama: hiyamas@nttdocomo.com
19 20 21	57	*Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp
22 23	58	
24 25	59	*Corresponding authors:
26 27 28	60	Junichi Sugawara: jsugawara@med.tohoku.ac.jp
29 30	61	and Masao Nagasaki: nagasaki@megabank.tohoku.ac.jp
31 32	62	Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi,
33 34 35	63	Aoba-ku, 980-8573, Sendai, Japan. Phone: +81-22-273-6283
35 36 37	64	
38 39	65	Word count: 5305 words
40 41	66	Key words: lifelog, multi-omics analysis, prediction, complicated pregnancy
42 43 44	67	
45 46	68	
47 48	69	
49 50 51	70	
52 53	71	
54 55	72	
56 57	73	
58 59 60	74	
		3

Page 5 of 43

Abstract

BMJ Open

ed,	MJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.
sing	rom http://bmjopen.bmj.com/ on June 11, 2025 at Agen (ABES). and data mining, AI training, and similar technologies.
y of	mj.com raining
d	/ on Ju ا, and s
ody	ıne 11, similar
me	2025 at technol
ber	Agence Biblic ogies.
	ographique d
4	le I Enseignemer
	t

ΒM

76	Purpose: A prospective cohort study for pregnant women, the Maternity Log
77	study (MLOG), was designed to construct a time-course high-resolution
78	reference catalog of bioinformatic data in pregnancy and explore the
79	associations between genomic and environmental factors and the onset of
80	pregnancy complications, such as hypertensive disorders of pregnancy,
81	gestational diabetes mellitus, and preterm labor, using continuous lifestyle
82	monitoring combined with multi-omics data on the genome, transcriptome,
83	proteome, metabolome, and microbiome.
84	Participants: Pregnant women were recruited at the timing of first routine
85	antenatal visits at Tohoku University Hospital, Sendai, Japan between
86	September 2015 and November 2016. Of the eligible women who were invited,
87	65.4% agreed to participate, and a total of 302 women were enrolled. The
88	inclusion criteria were age \geq 20 years and the ability to access the internet using
89	a smartphone in the Japanese language.
90	Findings to date:
91	Study participants uploaded daily general health information including quality of
92	sleep, condition of bowel movements, and the presence of nausea, pain, and
93	uterine contractions. Participants also collected physiologic data, such as body

94 weight, blood pressure, heart rate, and body temperature, using multiple home

- 95 healthcare devices. The mean upload rate for each lifelog item was ranging
- 96 from 67.4 % (fetal movement) to 85.3% (physical activity) and the total number
- 97 of data points was over 6 million. Biospecimens, including maternal plasma,
- 98 serum, urine, saliva, dental plaque, and cord blood, were collected for multi-
- 99 omics analysis.

100	Future plans:
101	Lifelog and multi-omics data will be used to construct a time-course high-
102	resolution reference catalog of pregnancy. The reference catalog will allow us to
103	discover relationships among multi-dimensional phenotypes and novel risk
104	markers in pregnancy for the future personalized early prediction of pregnancy
105	complications.
106	
107	Strengths and limitations of this study:
108	This is the first study designed to collect longitudinal lifelog information
109	through healthcare devices, self-administered questionnaires using
110	smartphones, and varieties of biospecimens throughout pregnancy.
111	 Longitudinal, continuous, individual lifelog data with a high acquisition rate
112	will enable us to assess dynamic physiological changes throughout
113	pregnancy.
114	 Multi-omics data will make it possible to understand the complex
115	mechanisms of multifactorial pregnancy-related diseases.
116	Potential limitations are the limited sample size and participant recruitment
117	only at a tertiary hospital for high-risk populations.
118	 Inclusion criteria of the present study limited the eligibility to pregnant
119	women with age >20 years and the ability to access the internet using a
120	smartphone.
121	
122	
123	
124	
	Ę
	101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122

1 2

hospital for high-risk populations.	
a of the present study limited the eligibility to pregnant	
e >20 years and the ability to access the internet using a	

5

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Page 7 of 43

125 INTRODUCTION	125	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 multiomics analyses and objective data on environmental factors, including daily lifelog data, in pregnant women. This study may demonstrate correlations between specific lifelog patterns and pregnancy related physiological changes, such as blood pressure, gestational weight gain, and onset of obstetric diseases. Furthermore, studies on associations among lifelog patterns, plasma and urine metabolomes, transcriptomes, and genomic variations may reveal relationships among 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.

 ABMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement

 Superieur (ABES) .

 Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

 P

 P

BMJ Open

150 COHORT DESCRIPTION

151 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 3generation 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

Page 9 of 43

1 2		
3 4	175	and NTT DoCoMo, Inc. (Tokyo, Japan).
5 6 7	176	
8 9	177	Patient and public involvement
10 11	178	Patients or the public were not directly involved in the development of the
12 13 14	179	research question or the design of the study. The main results will be made
14 15 16	180	available in the public domain.
17 18	181	
19 20	182	Participants
21 22 23	183	Participants were recruited at a first routine antenatal visit at Tohoku University
24 25	184	Hospital, Sendai, Japan between September 2015 and November 2016. A
26 27	185	flowchart of the recruitment process is shown in Figure 1. GMRCs at Tohoku
28 29 30 31 32	186	University Hospital approached eligible pregnant women for TMM BirThree
	187	Cohort Study (n= 631), and patients who already agreed to participate in TMM
33 34	188	BirThree Cohort Study (n=513) were assessed for eligibility for the MLOG study.
35 36 27	189	Finally, 462 pregnant women were asked to provide informed consent for the
37 38 39	190	MLOG study. A total of 302 women were enrolled. The inclusion criteria were
40 41	191	the age \geq 20 years and the ability to access the internet using a smartphone in
42 43 44 45 46	192	the Japanese language. Participants were excluded after enrollment if
	193	termination of pregnancy, abortion, or transfer to another institution for
47 48	194	emergency care occurred before delivery, or if they withdrew consent for any
49 50	195	reason.
51 52 53	196	
55 55	197	Outline of study protocol
56 57	198	The study protocol consisted of blood and urine sampling, saliva and dental
58 59	199	plaque sampling, self-administered daily lifelog data collection, and data upload
60		8

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.

208 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.

BMJ Open

2 3		
4 5	225	Saliva and dental plaque sampling
5 6 7	226	Samples of saliva and dental plaque were collected 3 times from each
8 9 10	227	participant, at the same time points as blood collection. Approximately 3 mL of
10 11	228	saliva was collected using a 50-mL conical centrifuge tube (Corning, Inc.,
12 13 14	229	Corning, NY, USA) and stored at -80°C until analysis. Dental plaque was
15 16	230	sampled by brushing, suspended in 0.5 mL of Tris-EDTA (10 mM Tris, 1 mM
17 18	231	EDTA; pH, 8.0), and immediately stored at -80°C until the time of sample
19 20	232	processing.
21 22 23	233	
24 25	234	Lifelog data collection
26 27	235	Based on previous publications on the utility for risk assessment of pregnancy-
28 29 30 31 32	236	related diseases, we selected several lifelog parameters to employ in this study,
	237	<i>i.e.</i> , body temperature [10], home blood pressure [11], body weight [12], and
33 34	238	physical activity (calorie expenditure) [13], as well as self-administered
35 36 37 38 39	239	information such as sleep quality [14], condition of stool [15], severity of nausea
	240	[16], fetal movement [17], severity of pain [18], uterine contractions [19], and
40 41	241	palpitations [20]. Body temperature, home blood pressure, body weight, and
42 43	242	physical activity were uploaded from multiple healthcare devices through a
44 45 46 47 48	243	smartphone. The self-administered information described above was input
	244	manually on mobile applications created for this study.
49 50	245	Data collection was started after obtaining informed consent and after giving
51 52 53	246	detailed instructions for the use of the healthcare devices. These applications
54 55	247	tracked quality of sleep; condition of stool using the Bristol Scale [21-23];
56 57	248	severity of nausea using the Pregnancy-Unique Quantification of Emesis and
58 59	249	nausea (PUQE) score [24,25]; headache, toothache, lumbago, and upper and
60		10

2 3		
4 5	250	lower abdominal pain using a numerical rating scale (NRS) score; the number
6 7 8 9	251	of perceived uterine contractions; palpitations; and fetal movement using a
	252	modified count-to-10 fetal movement chart [26,27].
10 11	253	Sleep quality was evaluated by the wakeup time, bedtime, sleep satisfaction
12 13 14	254	(ranked from satisfied to poor using a numeric scale of 0-4), and the number of
15 16	255	nocturnal awakenings (0-6).
17 18	256	The Bristol stool form scale was originally developed to assess constipation
19 20 21	257	and diarrhea [21, 22], and its use has been spread widely to evaluate functional
21 22 23	258	bowel disorders [22]. Using the Bristol scale, stool is classified into 7 types
24 25	259	according to cohesion and surface cracking [21, 22].
26 27	260	The PUQE score [24, 25] was developed to estimate the severity of nausea
28 29 30 31 32 33 34 35 36 37	261	and vomiting in pregnancy and quantifies the number of daily vomiting and
	262	retching episodes and the length of nausea in hours (over the preceding 12 h).
	263	The total score ranges from 3 (no symptoms) to 15, and higher scores are
	264	correlated with increasing severity of nausea and vomiting [24, 25].
38 39	265	In the NRS score for headache, toothache, lumbago, and upper and lower
40 41	266	abdominal pain, the total score ranges from 0 (no pain) to 10 (maximum ever
42 43 44	267	experienced).
44 45 46	268	Uterine contractions and palpitations were evaluated using definitions
40 47 48	269	determined for the current study. Uterine contractions were assessed using the
49 50	270	number of perceived contractions per day, ranging from 0 to more than 5. The
51 52 53	271	count-to-10 method was originally developed to assess fetal well-being by
54 55	272	recording the time, in minutes, required to count 10 fetal movements [26]. More
56 57	273	recently, a modified count-to-10 method has been proposed: pregnant women
58 59 60	274	are advised to start counting when they feel the first movement, then record the
00		11

Page 13 of 43

BMJ Open

	Superieur (ABES) . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.	BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement
--	--	---

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

 ABMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement

 Superieur (ABES) .

 Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

 Page

 Page

2 3	
4	
5	
6	
6 7	
8	
9	
10	
11	
12	
13	
14 15	
15 16	
16 17	
18	
19	
21	
20 21 22 23	
23	
24 25	
25	
26 27	
27	
29	
30	
31	
32	
33	
34	
35	
36	
37 38	
30 39	
40	
41	
42	
43	
44	
45	
46	
47 48	
40 49	
50	
51	
52	
53	
54	
55	
56 57	
57 58	
58 59	
60	
-	

1 2

Hospital. Epidemiological data, including extensive questionnaire surveys by
TMM BirThree Cohort Study can be obtained from the ToMMo integrated
biobank [8].

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.

308 Data handling was strictly regulated under HIPAA (Health Insurance Portability

and Accountability Act of 1996, United States Security and Privacy Rules) [28,

310 29] and the Act on the Protection of Personal Information [30]. Security control

311 at our facility has been described previously [31].

312

303

304

Database

313 Omics analysis

314 Whole-genome sequencing

315 To minimize amplification bias, we adopted a PCR-free library preparation 316 method. After performing library quality control using the quantitative MiSeq 317 method [32], libraries were sequenced on HiSeg 2500 Sequencing System 318 (Illumina, Inc., San Diego, CA, USA) to generate 259-bp, paired-end reads. We 319 generated the sequencing data at over 12.5x coverage on average, and we 320 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 321 322 called across all samples using Genome Analysis Tool Kit's HaplotypeCaller 323 (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 324

1 2		
3 4	325	genome was GRCh37/hg19 with the decoy sequence (hs37d5) and NC_007605
5 6 7	326	(Human Gamma Herpesvirus 4). The complete fasta file named
, 8 9	327	hg19_tommo_v2.fa is available from iJGVD website
10 11	328	(http://ijgvd.megabank.tohoku.ac.jp) [33]. For the quality assurance, we have
12 13 14	329	checked the ratio of the bases with the phred quality score over 30, the total
15 16	330	variant numbers in each chromosome, and the ratio of transitions to
17 18	331	transversions for a pair of sequences.
19 20 21	332	
22 22 23	333	Transcriptome
24 25	334	Whole blood was collected using the $PAXgene$ RNA tube, which is widely
26 27 28	335	used for transcriptome analysis. After storage at -80°C, total RNA was purified
28 29 30	336	with PAXgene Blood RNA Kit $^{ m I\!R}$ (Qiagen, Venlo, The Netherlands) using
31 32	337	QiaSymphony ${ m I\!R}$ (Qiagen). Total RNA was reverse-transcribed using an oligo-dT
33 34	338	primer. We used TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc.)
35 36 37	339	for library preparation for sequencing with HiSeq 2500 Sequencing System. For
38 39	340	the quality assurance, we randomly selected 11 samples in one batch (usually
40 41	341	48 samples) and checked an RNA integrity number (RIN) (or an RIN equivalent)
42 43 44	342	using BioAnalyzer ${ m I\!R}$ or Tape Station ${ m I\!R}$ (both from Agilent Technologies, Santa
45 46	343	Clara, CA, USA). The batch with RIN (or an RIN equivalent) higher than 7.0 for
47 48	344	all tested samples was used for the downstream analysis. The minimum
49 50 51	345	threshold for the total sequence reads for each sample was set to thirty millions.
52 53	346	For computing a series of quality control metrics for RNA-seq data, RNA-SeQC
54 55	347	was used to check the quality of sequence reads [34].
56 57	348	
58 59 60	349	Plasma and urine metabolome

350	Nuclear magnetic resonance (NMR) spectroscopy
351	All NMR measurements for metabolome analysis were conducted at 298 K on a
352	Bruker Avance 600 MHz spectrometer equipped with a SampleJet sample
353	changer (Bruker Corp., Billerica, MA, USA) [35]. Standard 1-dimensional
354	nuclear Overhauser enhancement spectroscopy (1D NOESY) and Carr-Purcell-
355	Meiboom-Gill (CPMG) spectra were obtained for each plasma or urine sample.
356	All spectra for plasma or urine samples were acquired using 16 scans and 32 k
357	of complex data points. All data were analyzed using the TopSpin 3.5 (Bruker
358	Corp.) and Chenomx NMR Suite 8.2 (Chenomx Inc., Edmonton, Alberta,
359	Canada) programs. All spectra were referenced to an internal standard (DSS-
360	d6). As necessary, those spectra were aligned using hierarchical cluster-based
361	peak alignment method, which is implemented as an R package called "speaq"
362	[36].
363	
364	Gas chromatography-tandem mass spectrometry (GC-MS/MS)
365	Sample preparation for plasma and urine (50 μL each) was performed using a
366	Microlab STARIet robot system (Hamilton, Reno, NV, USA) followed by the
367	methods previously reported by Nishiumi [37, 38]. The resulting deproteinized
368	and derivatized supernatant (1 μL) was subjected to GC-MS/MS, performed on
369	a GC-MS TQ-8040 system (Shimadzu Corp., Kyoto, Japan). The compound
370	separation was performed using a fused silica capillary column (BPX-5; 30 m $ imes$
371	0.25 mm inner diameter; film thickness, 0.25 μ m; Shimadzu Corp, Kyoto,
372	Japan). Metabolite detection was performed using Smart Metabolites Database
373	(Shimadzu Corp.) that contained the relevant multiple reaction monitoring
374	(MRM) method file and data regarding the GC analytical conditions, MRM

375	parameters, and retention index employed for the metabolite measurement. The
376	database used in this study included data on 475 peaks from 334 metabolites.
377	All peaks of metabolites detected from each sample was annotated and
378	analyzed using Traverse $MS^{I\!\!R}$ (Reifycs Inc., Tokyo, Japan). Then, two types of
379	normalization were performed to these annotated metabolites. The first
380	normalization was performed using the peak of 2-isopropylmalic acid as an
381	internal standard which was added to each sample before analysis with GC-
382	MS/MS. Then the second normalization was performed using quality control
383	(QC) samples which were injected after every 12 study samples according to
384	the RQC normalization methods [39]. Normalized values of each metabolite in
385	the QC samples were assessed by calculating coefficients of variation (CVs),
386	and metabolites with CVs over 20% were eliminated.

388 Oral Microbiome

Analysis of oral microbiome was conducted by previously reported protocols [40]. In brief, saliva was collected in a 50-mL tube. Dental plague 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.

409 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.

- - 436 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. Page 21 of 43

BMJ Open

2		
3 4 5	475	Standard analyses would be also applicable for the selected variables and
5 6 7	476	extracted factors (features). The association of outcomes with each feature will
8 9	477	be analyzed using statistical hypothesis tests such as Welch's t-test, Fisher's
10 11	478	exact test, the Chi-square test, and others as appropriate. Multiple logistic
12 13 14	479	regression modeling will be used to adjust for confounders and to assess
15 16	480	whether each feature or combination of features can be used to predict
17 18	481	outcomes. Stepwise selection algorithms or regularized algorithms (e.g.,
19 20 21	482	LASSO, ridge regression, or elastic net) will be used to select the optimal
22 22 23	483	number of contributing features that maximize the predictive power using the
24 25	484	leave-1-out cross validation or K-fold cross validation methods.
26 27	485	Individual genetic features may have an effect on outcomes; therefore, some
28 29 30	486	aggregated genetic risk score should be included in the prediction model. For
31 32	487	example, SNVs, including rare variants in or around a chromosome region of a
33 34	488	known or estimated risk gene, could be aggregated by considering their impacts
35 36 37	489	on biological function of the gene or their minor allele frequencies in the
38 39	490	population. However, this study is limited in the number of study participants,
40 41	491	and the aggregated risk score might therefore contribute only slightly to the
42 43 44	492	predictive power. To create a more reliable risk score, the estimates from other
45 46	493	large-scale cohort data using polygenic score tools, e.g., PRSice [55], could be
47 48	494	used for this study.
49 50	495	
51 52 53	496	FINDINGS TO DATE
54 55	497	Clinical background
56 57	498	A total of 302 women were enrolled, and the mean gestational weeks of
58 59 60	499	recruitment was 16.4 \pm 4.9 weeks (mean \pm SD). A total of 285 participants have
00		20

1 2	
2 3	
4	
5	
6	
7	
8	
9	
10	
11	
12 13	
14	
15	
16	
17	
18	
19 20	
20	
22	
23	
24	
25	
26 27	
28	
29	
30	
31	
32	
33 34	
35	
36	
37	
38	
39	
40 41	
41	
43	
44	
45	
46	
47 48	
40 49	
50	
51	
52	
53	
54 55	
55 56	
57	
58	
59	
60	

500	been followed up to delivery; their baseline clinical characteristics are described
501	in Table 1. The mean maternal age at delivery was 33.3 ± 4.9 years. As for
502	educational levels, 62% of the participants were high school graduates with or
503	without vocational college education, and 21% had a college degree. The
504	majority were employed (65%) in early pregnancy, and about 40% had a high
505	household income (over 6 million yen per year). Approximately 42% of the
506	participants were over 35 years of age, 51% were parous, and 22% were
507	overweight or obese by their prepregnancy body mass indices (BMI \ge 25
508	kg/m ²). Overall, 8.4% of the participants had HDP, and 5.6% underwent
509	spontaneous preterm birth. On average, infants were delivered at 38.0 ± 2.3
510	weeks of gestation with a mean birth weight of 2907 \pm 572 g. The rate of low
511	birth weight was 18%. Mean gestational weeks of the first and second blood
512	sampling were 17.0 \pm 5.0 and 27.5 \pm 2.5, respectively. The third blood sampling
513	was performed at 31.1 ± 3.0 days after delivery on average. The length of
514	enrollment ranged from 90 to 396 days with a mean of 216 \pm 61 days.
515 516	Table 4. Desticient characteristics

Characteristics	Value
Maternal (n = 285)	
\cdot 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)

1		
2		
3		
4 5	Vocational college	23 (28.4)
6	College degree and above	17 (21.0)
7		()
8	Others	1 (1.2)
9	Data not available	204
10		204
11		
12 13		
14	 Occupation (n=270) n (%) 	
15		
16	Housewife or unemployed	93 (34.4)
17	Employed	175 (64.8)
18		
19	Student	2 (0.7)
20 21		
22		
	Annual household income, yen (n=248) n (%)	
24		
25	< 2 million	17 (6.9)
26		50 (00 0)
27 28	2-4 million	59 (23.8)
29	4-6 million	73 (29.4)
30		
31	6-8 million	51 (20.6)
32	8-10 million	22 (8.9)
33		
34 35	> 10 million	26 (10.5)
36		
37		
38	・Parity, n (%)	
39	0	140 (49.1)
40	1	93 (32.6)
41 42	≥2	52 (18.2)
40		
44	 Prepregnancy BMI*, kg/m², mean (SD) 	22.7 (± 5.1)
	 Prepregnancy BMI, kg/m², n (%) 	
46	< 18.5	36 (12.6)
47	18.5-24.9	186 (65.3)
48	25.0-29.9	34 (11.9)
49 50	≥ 30.0	29 (10.2)
F 1		28.0.(+.2.2)
52	 Gestational weeks at delivery, mean (SD) 	38.0 (± 2.3)
53		
	• Mode of delivery, n (%)	
55	Noncesarean	179 (62.8)
56 57	Cesarean	106 (37.2)
57 58		
59	Pregnancy complication, n (%)	04 (0.4)
60	Hypertensive disorder of pregnancy	24 (8.4)

2	
3	
4	
5	
6	
5 6 7	
8	
9	
9 10	
11	
12 13	
13	
14	
14 15 16 17	
16	
17	
18	
19	
20	
 20 21 22 23 24 25 26 27 28 29 20 	
22	
25 24	
24 25	
25	
20	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41 42	
42 43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

1

Spontaneous preterm birth 16 (5.6) Neonatal (n = 300) 2907 (± 572) • Birth weight, g, mean (SD) 2907 (± 572) • Sex, n (%) 168 (56) Male 168 (56) Female 132 (44) • Low-birth weight (< 2500 g), n (%)</td> 54 (18) *BMI, body mass index

519 Data acquisition

517

518

The percentage of data uploads as of June 2017 was calculated for the 285 520 521 final study participants. For each lifelog item, the upload rate for each participant was calculated from the total number of days of actual uploads 522 divided by the number of days from enrollment to delivery. The mean upload 523 rate for each lifelog item was 85.3% (physical activity), 82.1% (body weight), 524 80.4% (body temperature), 78.0% (morning home blood pressure), 71.6% 525 (evening home blood pressure), 83.5% (sleep quality), 82.1% (condition of 526 stool, severity of pain, severity of nausea, uterine contractions, and 527 528 palpitations), and 67.4% (fetal movement) (Figure 3). 529

530 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.

Page 25 of 43

1

538	STRENGTHS AND LIMITATIONS
539	Herein, we have described the rationale, design, objective, data collection
540	methods, and interim results of the MLOG study. The study was launched in
) 1 541	September 2016, and baseline data collection ended in June 2017. A total of
2 3 542 4	285 participants uploaded lifelog data throughout pregnancy with a high data
5 5 5 5 5 43	acquisition rate and over 6 million total data points. Biospecimens for multi-
7 3 544	omics analysis were satisfactorily collected and all tracked by LIMS.
9) 545	There are three noteworthy features in the MLOG study. First, it is a
1 ² 546 3	prospective add-on cohort study based on TMM BirThree Cohort Study, with a
5 5 5 5 4 5 47	full series of epidemiological data and a highly structured follow-up system for
5 7 548	mothers, newborns, and families [8]. Second, we have successfully collected
3 9 549 0	longitudinal, continuous, individual lifelog data with a high acquisition rate,
1 2 550	which will enable us to assess dynamic changes in physiologic conditions
3 ₄ 551	throughout pregnancy. Third, multi-omics data will make it possible to fully
5 5 552 7	understand the complex mechanisms of multifactorial pregnancy-related
³ 553	diseases and to overcome the unpredictability of these complications.
) 1 554	Prediction models using clinical and epidemiological information and
2 3 555	circulating factors for pregnancy-related diseases have been developed
4 5 556	extensively [56], and risk-assessment approaches using clinical information
5 557	have also been developed [57, 58]. However, there is a lack of evidence for the
) 558	benefits of these predictive models for routine clinical use [59]. Once the
559	likelihood of a pregnancy-related disorder is estimated with high sensitivity and
3 4 560	specificity, evidence-based clinical interventions could reduce the rate of
5 7 561	maternal and neonatal morbidity and mortality [60]. Therefore, an early-
3 9 562	prediction algorithm that can be used with a high level of confidence is needed
)	24

 ABMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement

 Superieur (ABES) .

 Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

 Page

 Page

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

587 model for pregnancy-related diseases employing multi-omics and lifelog data.

Page 27 of 43

BMJ Open

1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19 20	
20	
21	
22	
23 24	
24 25	
25 26	
20 27	
27	
20 29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59 60	
60	

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.

601

602 Author affiliations

603 1 Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-

604 machi, Aoba-ku, Sendai, Miyagi, 980-8573, Japan.

- 605 2 Tohoku University Graduate School of Medicine, 1-1, Seiryo-machi, Aoba-ku,
- 606 Sendai, Miyagi, 980-8574, Japan.
- 607 3 Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka,
- 608 Kanagawa, Japan 239-8536.
- 609 4 Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-
- 610 ku, Sendai, Miyagi 981-8558, Japan.
- $\frac{1}{7}$ 611 5 Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku,
- ⁵⁹ 612 Hiroshima 732-0815, Japan.

BMJ Open: first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES)

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

> 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

1 2		
3 4	638	be accountable for all aspects of the work in ensuring that questions related to
5 6 7 8 9	639	the accuracy or integrity of any part of the work are appropriately investigated
	640	and resolved.
10 11 12	641	
12 13 14 15 16 17 18 19 20 21	642	Funding
	643	The present study was supported by NTT DoCoMo, Inc., with a collaborative
	644	research agreement between NTT DoCoMo and ToMMo. This work was
	645	supported in part by the Tohoku Medical Megabank Project from the Japan
22 23	646	Agency for Medical Research and Development (AMED) and the Ministry of
24 25	647	Education, Culture, Sports, Science and Technology (MEXT).
26 27 28	648	
28 29 30 31 32 33 34 35	649	Competing interests
	650	This study was funded by NTT DoCoMo, Inc.
	651	Daisuke Ochi, Takafumi Yamauchi, and Satoshi Hiyama are employees of NTT
36 37	652	DoCoMo, Inc. All other authors declare that they have no competing interests.
38 39	653	
40 41 42 43	654	Ethics approval and consent to participate
44 45	655	TMM BirThree Cohort Study was approved by the ethics committees of the
46 47	656	Tohoku University (authorization numbers, 2013-4-103 and 2017-4-010). The
48 49 50	657	MLOG study was approved by the ethics committees of the Graduate School of
50 51 52	658	Medicine (2014-1-704) and the Tohoku Medical Megabank Organization (2017-
53 54	659	1-085), Tohoku University. Written informed consent was obtained from all
55 56 57	660	participants.
58 59 60	661	20

1 2		
3 4	662	F
5 6 7	663	١
8 9	664	
10 11	665	۵
12 13 14	666	٧
14 15 16	667	٦
17 18	668	c
19 20	669	j
21 22 23	670	r
24 25	671	
26 27	672	
28 29 30	673	
31 32	674	
33 34	675	
35 36	676	
37 38 39	677	
40 41	678	
42 43	679	
44 45 46	680	
40 47 48	681	
49 50	682	
51 52	683	
53 54 55	684	
56 57	685	
58 59	686	
60		

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 sugawara@med.tohoku.ac.jp or Dr. Masao Nagasaki at nagasaki@megabank.tohoku.ac.jp. Currently, no additional data are available.

1 2		
3 4	687	
5 6 7	688	
8 9	689	
10 11 12	690	REFERENCES
12 13 14	691	1. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public
15 16	692	health perspective. Diabetes Care. 2007;30 Suppl 2: S141-6. doi:
17 18	693	10.2337/dc07-s206.
19 20 21	694	
22 23	695	2. Beck S, Wojdyla D, Say L, et al. The worldwide incidence of preterm birth: a
24 25	696	systematic review of maternal mortality and morbidity. Bull World Health Organ.
26 27 28	697	2010; 88:31-8. doi: 10.2471/BLT.08.062554.
29 30	698	
31 32	699	3. Duley L. The global impact of pre-eclampsia and eclampsia. Semin Perinatol.
33 34 35	700	2009; 33:130-7. doi: 10.1053/j.semperi.2009.02.010.
36 37	701	
38 39	702	4. Ananth CV, Keyes KM, Wapner RJ. Pre-eclampsia rates in the United
40 41 42	703	States, 1980-2010: age-period-cohort analysis. BMJ. 2013;347: f 6564. doi:
42 43 44	704	10.1136/bmj.f6564.
45 46	705	
47 48	706	5. Waken RJ, de Las Fuentes L, Rao DC. A Review of the Genetics of
49 50 51	707	Hypertension with a Focus on Gene-Environment Interactions. Curr Hypertens
52 53	708	Rep. 2017; 19:23. doi: 10.1007/s11906-017-0718-1.
54 55	709	
56 57 58	710	6. Ward K, Lindheimer MD. Genetic factors in the etiology of preeclampsia /
59 60	711	eclampsia. In: Chesley's Hypertensive Disorders in pregnancy. London:

1 2		
3 4 5	712	Elsevier; 2990: 51-72.
6 7	713	
8 9 10	714	7. Li X, Dunn J, Salins D, Zhou G, et al. Digital Health: Tracking Physiomes and
10 11	715	Activity Using Wearable Biosensors Reveals Useful Health Related Information.
12 13 14	716	PLoS Biol. 2017; 15: e2001402. doi: 10.1371/journal.pbio.2001402.
15 16	717	
17 18	718	8. Kuriyama S, Yaegashi N, Nagami F, et al. The Tohoku Medical Megabank
19 20 21	719	Project: Design and Mission. J Epidemiol. 2016; 26:493-511. doi:
21 22 23	720	10.2188/jea.JE20150268.
24 25	721	
26 27 28 29 30 31 32	722	9. Japan Society of Obstetrics and Gynecology, Guideline for Obstetrical
	723	Practice in Japan, Japan Society of Obstetrics and Gynecology, Tokyo, Japan,
	724	pp. 1–4, 2017 [in Japanese].
33 34	725	
35 36 37	726	10. Hartgill TW, Bergersen TK, Pirhonen J. Core body temperature and the
38 39	727	thermoneutral zone: a longitudinal study of normal human pregnancy. Acta
40 41	728	Physiol (Oxf). 2011; 201: 467-74. doi: 10.1111/j.1748-1716.2010.02228.x.
42 43	729	
44 45 46	730	11. Metoki H, Ohkubo T, Watanabe Y, et al. Seasonal trends of blood pressure
47 48	731	during pregnancy in Japan: the babies and their parents' longitudinal
49 50	732	observation in Suzuki Memorial Hospital in Intrauterine Period study. J
51 52 53	733	Hypertens. 2008; 26: 2406-13. doi: 10.1097/HJH.0b013e32831364a7.
54 55	734	
56 57	735	12. Haugen M, Brantsæter AL, Winkvist A, et al. Associations of pre-pregnancy
58 59	736	body mass index and gestational weight gain with pregnancy outcome and
60		31

2 3		
5 4 5	737	postpartum weight retention: a prospective observational cohort study. BMC
6 7	738	Pregnancy Childbirth. 2014 Jun 11; 14: 201. doi: 10.1186/1471-2393-14-201.
8 9	739	
10 11	740	13. Sorensen TK, Williams MA, Lee IM, et al. Recreational physical activity
12 13 14	741	during pregnancy and risk of preeclampsia. Hypertension. 2003 Jun; 41:1273-
15 16	742	80. doi: 10.1161/01.HYP.0000072270.82815.91
17 18	743	
19 20 21	744	14. Reutrakul S, Zaidi N, Wroblewski K, et al. Sleep disturbances and their
21 22 23	745	relationship to glucose tolerance in pregnancy. Diabetes Care. 2011; 34: 2454-
24 25	746	7. doi: 10.2337/dc11-0780.
26 27	747	
28 29 30	748	15. Cornish J, Tan E, Teare J, et al. A meta-analysis on the influence of
31 32	749	inflammatory bowel disease on pregnancy. Gut. 2007; 56: 830-7. doi:
33 34	750	10.1136/gut.2006.108324.
35 36	751	
37 38 39	752	16. Huxley RR. Nausea and vomiting in early pregnancy: its role in placental
40 41	753	development. Obstet Gynecol. 2000; 95:779-82.
42 43	754	
44 45 46	755	17. Holm Tveit JV, Saastad E, Stray-Pedersen B, et al. Maternal characteristics
40 47 48	756	and pregnancy outcomes in women presenting with decreased fetal movements
49 50	757	in late pregnancy. Acta Obstet Gynecol Scand. 2009; 88: 1345-51. doi:
51 52	758	10.3109/00016340903348375.
53 54 55	759	
56 57	760	18. Facchinetti F, Allais G, D'Amico R, et al. The relationship between
58 59	761	headache and preeclampsia: a case-control study. Eur J Obstet Gynecol
60		32

1 2			
3 4 5	762	Reprod Biol. 2005; 121: 143-8. doi: 10.1016/j.ejogrb.2004.12.020.	
5 6 7 8 9	763		
	764	19. lams JD, Newman RB, Thom EA, et al; National Institute of Child Health	
10 11	765	and Human Development Network of Maternal-Fetal Medicine Units. Frequen	су
12 13	766	of uterine contractions and the risk of spontaneous preterm delivery. N Engl J	
14 15 16	767	Med. 2002; 346: 250-5. doi: 10.1056/NEJMoa002868	
17 18	768		
19 20	769	20. Abbas AE, Lester SJ, Connolly H. Pregnancy and the cardiovascular	
21 22 23	770	system. Int J Cardiol. 2005; 98: 179-89. doi: 10.1016/j.ijcard.2003.10.028.	
23 24 25	771		
26 27	772	21. Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal tran	sit
28 29 30 31 32	773	time. Scand J Gastroenterol. 1997; 32: 920-4.	
	774	doi: 10.3109/00365529709011203	
33 34	775		
35 36 37 38 39	776	22. Riegler G, Esposito I. Bristol scale stool form. A still valid help in medical	
	777	practice and clinical research. Tech Coloproctol 2001; 5: 163–4. doi:	
40 41	778	10.1007/s101510100019	
42 43	779		
44 45 46 47 48 49 50 51 52 52	780	23. Longstreth GF, Thompson WG, Chey WD, et al. Functional Bowel	
	781	Disorders. Gastroenterology 2006; 130: 1480–91. doi:	
	782	10.1053/j.gastro.2005.11.061	
	783		
53 54 55	784	24. Koren G, Boskovic R, Hard M, et al. Motherisk-PUQE (pregnancy-unique	
56 57	785	quantification of emesis and nausea) scoring system for nausea and vomiting	of
58 59	786	pregnancy. Am J Obstet Gynecol. 2002;186: S228-31.	
60			33

Page 35 of 43

1

2		
3 4	787	
5 6 7	788	25. Koren G, Piwko C, Ahn E, et al. Validation studies of the Pregnancy Unique-
8 9	789	Quantification of Emesis (PUQE) scores. J Obstet Gynaecol. 2005; 25: 241-4.
10 11	790	doi: 10.1080/01443610500060651
12 13 14	791	
15 16	792	26. Pearson JF, Weaver JB. Fetal activity and fetal wellbeing: an evaluation. Br
17 18	793	Med J 1976; 1:1305–7.
19 20 21	794	
22 23	795	27. Winje BA, Saastad E, Gunnes N, et al. Analysis of 'count-to-ten' fetal
24 25	796	movement charts: a prospective cohort study. BJOG. 2011;118: 1229-38. doi:
26 27	797	10.1111/j.1471-0528.2011.02993.x
28 29 30	798	
31 32	799	28. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach
33 34	800	Notification rules under the Health Information Technology for Economic and
35 36 37	801	Clinical Health Act and the Genetic Information Nondiscrimination Act; other
38 39	802	modifications to the HIPAA rules. Fed Regist. 2013; 78: 5565-702.
40 41	803	
42 43 44	804	29. Health Insurance Portability and Accountability Act of 1996. Public Law 104-
44 45 46	805	191. US Statut Large. 1996; 110:1936-2103.
47 48	806	
49 50	807	30. Amended Act on the Protection of Personal Information.
51 52 53	808	https://www.ppc.go.jp/files/pdf/Act_on_the_Protection_of_Personal_Information.
54 55	809	pdf.
56 57	810	
58 59 60	811	31. Takai-Igarashi T, Kinoshita K, Nagasaki M, et al. Security controls in an
00		34

2		
3 4 5	812	integrated Biobank to protect privacy in data sharing: rationale and study
6 7 8 9	813	design. BMC Med Inform Decis Mak. 2017; 17:100. doi: 10.1186/s12911-017-
	814	0494-5
10 11	815	
12 13 14	816	32. Katsuoka F, Yokozawa J, Tsuda K, et al. An efficient quantitation method of
15 16	817	next-generation sequencing libraries by using MiSeq sequencer. Anal Biochem.
17 18	818	2014; 466: 27-9. doi: 10.1016/j.ab.2014.08.015
19 20 21	819	
21 22 23	820	33.Yamaguchi-Kabata Y, et al, Nariai N, Kawai Y, et al. iJGVD: an
24 25 26 27 28 29 30 31 32 33 34 35 36 37	821	integrative Japanese genome variation database based on whole-genome
	822	sequencing. Hum Genome Var. 2015; 2:15050. doi: 10.1038/hgv.2015.50.
	823	https://www.ncbi.nlm.nih.gov/pubmed/27081555.
	824	
	825	34. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics
	826	for quality control and process optimization. Bioinformatics. 2012; 28: 1530-
38 39	827	1532. doi: 10.1093/bioinformatics/bts196.
40 41	828	
42 43 44	829	35. Koshiba S, Motoike I, Kojima K, et al. The structural origin of metabolic
45 46	830	quantitative diversity. Sci Rep. 2016; 6: 31463. doi: 10.1038/srep31463
47 48	831	
49 50 51	832	36. Vu TN, Valkenborg D, Smets K, et al. An integrated workflow for robust
52 53	833	alignment and simplified quantitative analysis of NMR spectrometry data. BMC
54 55	834	Bioinformatics. 2011; 12: 405. doi: 10.1186/1471-2105-12-405.
56 57 58	835	
58 59 60	836	37. Nishiumi S, Kobayashi T, Ikeda A, et al. A novel serum metabolomics-based

1 2		
3 4	837	diagnostic approach for colorectal cancer. PLoS One. 2012; 7: e40459. doi:
5 6 7	838	10.1371/journal.pone.0040459.
, 8 9	839	
10 11 12	840	38. Nishiumi S, Kobayashi T, Kawana S, et al. Investigations in the possibility of
12 13 14	841	early detection of colorectal cancer by gas chromatography/triple-quadrupole
15 16	842	mass spectrometry. Oncotarget. 2017; 8, 17115-17126. doi:
17 18	843	10.18632/oncotarget.15081.
19 20 21	844	
21 22 23	845	39. Saigusa D, Okamura Y, Motoike IN, et al. Establishment of Protocols for
24 25 26 27 28 29 30 31 32 33 34 35	846	Global Metabolomics by LC-MS for Biomarker Discovery. PLoS One.
	847	2016;11(8): e0160555. doi: 10.1371/journal.pone.0160555.
	848	
	849	40. Sato Y, Yamagishi J, Yamashita R, et al. Inter-Individual Differences in the
	850	Oral Bacteriome Are Greater than Intra-Day Fluctuations in Individuals. PLoS
35 36 37	851	One. 2015;10: e0131607. doi: 10.1371/journal.pone.0131607.
38 39	852	
40 41	853	41. Brown MA, Magee LA, Kenny LC, et al. Hypertensive Disorders of
42 43 44	854	Pregnancy: ISSHP Classification, Diagnosis, and Management
45 46	855	Recommendations for International Practice. Hypertension. 2018; 72: 24–43.
47 48	856	doi: 10.1161/HYPERTENSIONAHA.117.10803.
49 50 51	857	
52 53	858	42. Watanabe K, Naruse K, Tanaka K, et al. Outline of definition and
54 55	859	classification of pregnancy induced hypertension (PIH). Hypertens Res
56 57 58	860	Pregnancy 2013; 1: 3–4.
58 59 60	861	

2	
3 4	862
5	
6 7	863
8	864
9 10	
11	865
12 13	866
14	000
15 16	867
17	868
18 19	000
20	869
21 22	870
23	070
24 25	871
26	872
27 28	072
29 30	873
31	874
32 33	074
34	875
35 36	876
37	070
38 39	877
40	878
41 42	070
43	879
44 45	880
46	000
47 48	881
49	882
50 51	002
52	883
53 54	001
55	884
56 57	885
58	000
59	886

1

43. IADPSG Consensus Panel: International association of diabetes and

863 pregnancy study groups recommendations on the diagnosis and classification

864 of hyperglycemia in pregnancy. Diabetes Care 2010; 33: 676-682. doi:

865 10.2337/dc09-1848.

44. Box GEP, Jenkins GM, Reinsel GC. Time series Analysis: Forecasting andControl. 5th ed. New Jersey: Wiley; 2015.

45. Brandt PT, Williams JT. Multiple Time Series Models, Thousand Oaks, CA:

871 Sage Publications, 2007.

46. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data,

874 Journal of Statistical Software, 45 (7) 2011.

876 47. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by

877 surrogate variable analysis. PLoS Genet. 2007; 3: 1724-35. doi:

878 10.1371/journal.pgen.0030161

48. Stegle O, Parts L, Piipari M, et al. Using probabilistic estimation of

881 expression residuals (PEER) to obtain increased power and interpretability of

882 gene expression analyses. Nat Protoc. 2012; 7: 500-7. doi:

883 10.1038/nprot.2011.457.

49. GTEx Consortium, Battle A, Brown CD, et al. Genetic effects on gene

⁵⁹ 886 expression across human tissues. Nature. 2017; 550: 204-213. doi:

1 2		
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32	887	10.1038/nature24277.
	888	
	889	50. Polgreen PM, Yang M, Kuntz JL, et al. Using oral vancomycin prescriptions
	890	as a proxy measure for Clostridium difficile infections: a spatial and time series
	891	analysis. Infect Control Hosp Epidemiol. 2011; 32: 723-6. doi: 10.1086/660858.
	892	
	893	51. McDowell IC, Manandhar D, Vockley CM, et al. Clustering gene expression
	894	time series data using an infinite Gaussian process mixture model. PLoS
	895	Comput Biol. 2018; 14: e1005896. doi: 10.1371/journal.pcbi.1005896.
	896	
	897	52. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of
	898	Structured Time-Series. IEEE Trans Pattern Anal Mach Intell. 2015; 37: 383-93.
	899	doi: 10.1109/TPAMI.2014.2318711.
33 34	900	
35 36 37	901	53. Rohart F, Gautier B, Singh A et al. mixOmics: An R package for 'omics
37 38 39	902	feature selection and multiple data integration. PLoS Comput Biol. 2017;13:
40 41	903	e1005752. doi: 10.1371/journal.pcbi.1005752.
42 43 44	904	
45 46	905	54. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a
47 48	906	framework for unsupervised integration of multi-omics data sets. Mol Syst Biol.
49 50 51	907	2018; 14: e8124. doi: 10.15252/msb.20178124.
52 53	908	
54 55	909	55. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software.
56 57 58 59 60	910	Bioinformatics, 2015; 31:1466-8. doi: 10.1093/bioinformatics/btu848.
	911	
		38

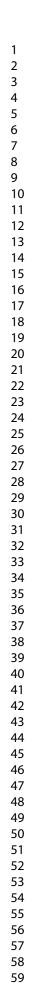
912	56. Wax JR, Cartin A, Pinette MG. Biophysical and Biochemical Screening for
913	the Risk of Preterm Labor: An Update. Clin Lab Med. 2016; 36: 369-83. doi:
914	10.1016/j.cll.2016.01.019.
915	
916	57. Al-Rubaie Z, Askie LM, Ray JG, et al. The performance of risk prediction
917	models for pre-eclampsia 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
	913 914 915 916 917 918 920 921 922 923 924 923 924 925 926 927 928 926 927 928 929 928 929 920 921 923 923 923 923

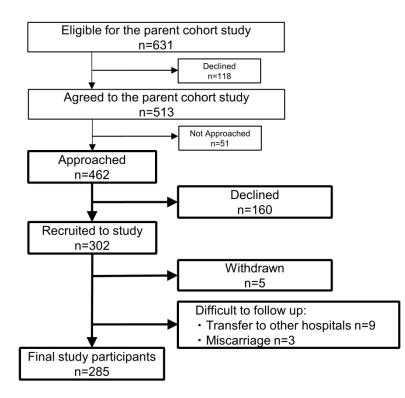
Page 41 of 43

1 2		
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 41 42 43 44 45 45 45 45 45 45 45 45 45	937	using personal, dense, dynamic data clouds. Nat Biotechnol. 2017; 35: 747-
	938	756. doi: 10.1038/nbt.3870.
	939	
	940	FIGURE TITLES AND LEGENDS
	941	
	942	Figure 1. Flowchart of Maternity Log Study (MLOG) participants
	943	
	944	Figure 2. Overview of the MLOG study protocol
	945	A: Participant timeline for the MLOG study.
	946	B: Physiologic information collected using healthcare devices. Specific
	947	measures were uploaded each day from the time of enrollment (solid horizontal
	948	lines). Participants had the option to continue uploading data until 180 days
	949	after delivery (dashed horizontal lines).
	950	C: Daily lifelogs of self-reported information using a smartphone application.
	951	Basic lifelog information was input manually from the time of enrollment (solid
	952	horizontal lines). Participants had the option to continue uploading data until
	953	180 days after delivery (dashed horizontal lines). Fetal movement and uterine
	954	contractions were recorded from 24 and 20 weeks of gestation, respectively.
	955	
46 47 48	956	Figure 3. Data acquisition rate
49 50	957	The mean data upload rate of specific measures was calculated from the total
51 52 53 54 55	958	number of days of actual uploads divided by the number of days from
	959	enrollment to delivery for each participant.
56 57		
58 59		
60		

43 perau first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de I Enseignement Superieur (ABES) . ge Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies. Pa

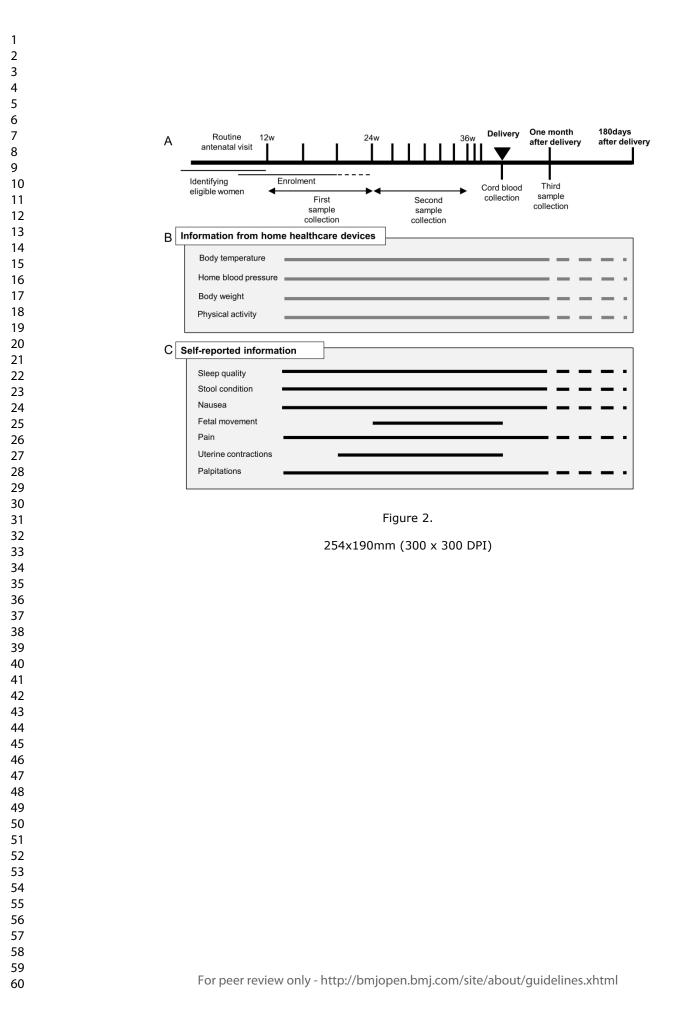
BMJ Open





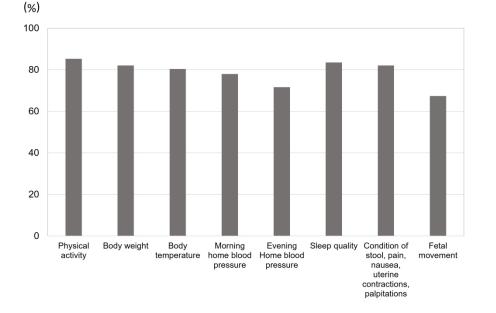


254x190mm (300 x 300 DPI)



pert first published as 10.1136/bmjopen-2018-025939 on 9 February 2019. Downloaded from http://bmjopen.bmj.com/ on June 11, 2025 at Agence Bibliographique de l Enseignement Superieur (ABES) . ge Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies. Pa

BMJ Open





254x190mm (300 x 300 DPI)

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml