National Centre for Research Methods working paper

Complex clinical data and Gestational Diabetes Mellitus

Jane Hirst, Huiqi Yvonne Lu, Lucy Mackillop and Yasmina Al Ghadban, University of Oxford

Ponnusamy Saravanan, University of Warwick

Sara White, King's College London

Economic and Social Research Council

Complex clinical data and Gestational Diabetes Mellitus

Professor Jane Hirst, Dr Lucy Mackillop, Ms Yasmina Al Ghadban
Nuffield Department of Women's & Reproductive Health
University of Oxford
jane.hirst@wrh.ox.ac.uk

Dr Huiqi Yvonne Lu
Department of Engineering Sciences
University of Oxford
yvonne.lu@eng.ox.ac.uk

Professor Ponnusamy Saravanan
Warwick Medical School
University of Warwick
P.Saravanan@warwick.ac.uk

Dr Sara White
Women and Children's Health
King's College London
sara.white@kcl.ac.uk

Abstract

This Innovation Forum brought together a multidisciplinary group of researchers, clinicians, data scientists, industry partners, NHS Digital, and others to discuss the opportunities and challenges for improving clinical care using complex clinical data. The workshop focussed on one specific clinical challenge, **Gestational Diabetes Mellitus (GDM)**, with consideration of other data science challenges and solutions in pregnancy care more widely and other clinical conditions.

The Forum was an opportunity to learn about specific data challenges researchers had experienced, learn from what has worked, and where further research is needed. We identified several opportunities for GDM data research in the UK, including the commissioning of the first national GDM audit and a need to focus on preventing type 2 diabetes after GDM in young women. The Forum identified six areas for future work and funding: (i) support for infrastructure to enable data science in this field; (ii) the need to map available data sources in the UK for pregnancy research; (iii) streamlined solutions for ethical approvals and regulatory support; (iv) improving data quality, linkage and access for researchers; (v) development of machine learning and statistical approaches; and (vi) the need to collaborate with clinicians, women and their families to make sure data science improves lives.

Keywords: Machine learning, statistics, gestational diabetes mellitus, clinical research, data science, NHS, pregnancy, women's health

Introduction

Complex clinical data are changing how healthcare is delivered. Increasing volumes of digital health data are accumulated in primary and secondary care, along with data from personal devices, sensors, third-party programs and 'Apps'. Whilst these data have enormous potential to improve health outcomes, and service delivery, only an estimated 2% of the 40 zettabytes of existing clinical data are used for research. Methodological barriers exist to accessing and using complex clinical data for research, making the process time-consuming, expensive, and challenging. Concerns exist around data privacy and protecting individual rights. There are also methodological challenges in combining and storing data from different platforms and locations, handling missing data, harmonising variables, and identifying the right opportunities where machine learning and other advanced statistical approaches will make a difference in how care is delivered.

Whilst these data challenges affect nearly all illnesses and health service delivery areas, this innovation forum considered these issues in the context of one specific clinical condition, gestational diabetes mellitus (GDM). The care of pregnant people with GDM is undergoing rapid changes driven by evolving technologies to monitor glucose and increasing prevalence, placing pressure on over-stretched NHS maternity services.

Overview of Gestational Diabetes Mellitus (GDM)

GDM is defined as glucose intolerance with first onset or recognition during pregnancy. It is estimated to affect approximately one in six pregnancies worldwide, making it the most common medical disorder of pregnancy. Due to increasing obesity and maternal age, prevalence is increasing worldwide, including in the UK [1]. This is placing increased demand on both maternity and diabetes clinical services.

GDM can lead to adverse pregnancy outcomes for both the mother and her baby, with a higher risk of preeclampsia, preterm birth and accelerated fetal growth, which can increase the risk of complications at birth, such as emergency caesarean section, shoulder dystocia and birth trauma. After birth, the baby is at risk of neonatal hypoglycaemia, which can be fatal or lead to permanent brain damage if severe and untreated [2]. Although glucose metabolism typically returns to normal after delivery, women who develop GDM are at extremely high risk of developing type 2 diabetes mellitus (T2DM), affecting up to 50% of GDM women within ten years of pregnancy. Women who develop GDM are also at increased risk of hypertension and cardiovascular disease [3–5], the leading causes of mortality in women globally. Additionally, GDM can impact the offspring, with the children experiencing an increased risk of T2DM, cardiovascular diseases and potentially even neurodevelopmental disorders such as ADHD and autism [6–15].

Despite being common, controversies remain around the identification and management of GDM during pregnancy and the screening and prevention of T2DM and other cardiometabolic complications after birth amongst women who are diagnosed with GDM. Currently, in the UK, NICE screening recommendations are for all pregnant women to be assessed for clinical risk factors for GDM early in pregnancy. Women with either a raised BMI (>30 kg/m^2), ethnic background with a higher risk of diabetes (South Asian, Middle Eastern, Black African/Black

Caribbean/Black British), family history of diabetes, or previous large baby (≥4.5 kg) are recommended to undergo a fasting, 2-hour, 75-gram oral glucose tolerance test at approximately 26-28 weeks of pregnancy. For those with a previous history of GDM, glucose testing earlier in pregnancy is recommended [16]. If the fasting or 2-hour glucose is above specific values, the diagnosis of GDM will be made.

There are several issues with this process. Screening based on clinical risk factors alone can miss approximately 30% of women with GDM. Additionally, with increasing numbers of women entering pregnancy with undiagnosed prediabetes or T2DM, diagnosis at 26-28 weeks may be too late to prevent complications [17,18]. There are also compliance challenges with GDM screening and testing, with some women missing testing altogether and others undergoing testing much later in pregnancy than recommended when it may be too late to prevent adverse outcomes such as macrosomia [19]. Blood glucose is a continuous measure. For decades, a debate has raged about appropriate diagnostic thresholds [20,21], whether the same thresholds should be used for all ethnic groups, at all gestations in pregnancy, and the cost-effectiveness of different screening approaches [22].

Once women are diagnosed with GDM, for most women, management follows a ‘one size fits all approach.’ There is little consideration of individual differences, yet it is known that certain groups face much higher risks during pregnancy, and GDM can have very different phenotypes. Management begins with education, dietary and lifestyle advice, and escalation to oral medications or insulin if insufficient to normalise blood glucose. Glucose levels are typically monitored by the pregnant woman at home using finger stick capillary testing. Whilst the NICE guidelines recommend all women with uncomplicated GDM be delivered by fourty weeks, and six days, the decision to or when and how to deliver is based mainly on clinical acumen considering the woman’s preferences, ultrasound estimation of fetal size and wellbeing, with a lack of evidence to guide birth choices.

In 2021, the James Lind Alliance (JLA) Priority Setting Partnerships (PSP) published the top 10 priorities for research in diabetes and pregnancy in the UK based on a consultation involving women, their support networks and healthcare professionals [23]. **The leading research priority was using diabetes technology to improve pregnancy, birth and mother and child health outcomes.** Diabetes technologies, such as digital glucose monitoring systems, continuous glucose monitoring and wearable devices, have the potential to provide new insights into GDM [24].

This innovation forum was held to discuss the specific challenges in accessing and using data to improve care for women with GDM, the digital data currently available for research in maternity care in the UK, data integration, standards and security and learning from case studies of other complex clinical data challenges.

Methodology

The NCRM Innovation Forum was an in-person meeting held on December 8-9th, 2022, at Jesus College, University of Oxford. The forum brought together obstetricians, endocrinologists, and other clinicians; researchers in maternal health, public health, medical statistics, and machine learning; data scientists, engineers, NHS Digital (Maternity Services Data Set); a representative from a funding body, and industry colleagues, with representing different expertise and levels of seniority (see participant list). The forum was themed around four topics (Appendix 1):

1. Routinely collected data in pregnancy research: Challenges and successes
2. Challenges accessing NHS and other data sources for research on women in pregnancy
3. Challenges combining different data platforms for research and clinical purposes
4. Methodological and analytic challenges: machine learning & classic statistical modelling

Participants were asked to share methodological challenges they had experienced in their own research, with discussions around the research priorities and opportunities to move forwards.

The workshop concluded with a final session where all participants discussed key priorities and gaps for data science to improve outcomes and clinical experiences for women with GDM.

Challenges and opportunities for data research in GDM

A new National Audit for GDM

The National Pregnancy in Diabetes Audit (NPID audit) was established over a decade ago in England. NPID is the largest population-level surveillance system globally on women with type 1 or 2 diabetes in pregnancy and maturity-onset diabetes of the young (MODY). NPID has enabled benchmarking across Trusts, monitoring trends in pregnancy preparedness and pregnancy, birth and perinatal outcomes. In contrast, no such national-level data exist for GDM, severely limiting the capacity to target policies and identify priority areas for research.

In 2021, Prof Eleanor Scott led a pilot GDM audit to determine the feasibility of a National GDM Audit for England using routinely collected data. Participating Trusts provided the NHS numbers of women diagnosed with GDM during a specified audit period. The NHS numbers were then linked to routinely collected data in the Maternity Services Data Set (MSDS) by NHS Digital. MSDS is a secondary-use dataset containing clinical and operational data for purposes other than direct patient care. MSDS version 2.0 was updated in 2018, mandating the submission of all maternity records in the scope of the dataset. The pilot audit sought to examine the feasibility of performing a national GDM audit using this dataset, with a focus on three questions:

(1) How many women have GDM?

(2) What are the birth outcomes of women with GDM?

(3) How many women diagnosed with GDM are screened after birth for T2DM?

Some key methodological learnings from the pilot audit were that whilst it was feasible to perform the audit across twenty Trusts, at the time of the pilot audit, large amounts of data were missing from MSDS. This likely reflected the relatively new requirement of mandatory reporting for the dataset, and the data completeness of MSDS continues to improve. The first National GDM Audit was commissioned in 2023. Hopefully, this will provide new data for analysis and highlight the importance of GDM to the public and health system. What is not counted doesn't count.

Lessons from research during the COVID-19 pandemic

COVID-19 necessitated a sudden and significant change in healthcare delivery for women with GDM. The 2-hour glucose tolerance test (OGTT) used to diagnose GDM was deemed potentially hazardous to women and health workers. Prof. Ponnusamy Saravanan led the RCOG emergency COVID guidance for GDM issued in April 2020 [25], changing the OGTT to a single blood test measuring HbA1c, a measure of average glucose control over the preceding weeks-months. A national group of clinicians and researchers led by Prof Rebecca Reynolds at the University of Edinburgh studied the effects of this change in diagnostic approach on outcomes for women with GDM and their babies.

This study's challenges represent barriers to multicentre data projects across the NHS. Whilst favourable ethical approval was obtained in Scotland, each Trust still needed site approvals before sharing anonymised data. Despite being a COVID-19 fast-track study, these approvals took many months. Data extraction was also challenging, as data were captured using 14 different platforms. Each trust had a different approach with varying levels of granularity and different views on what they considered identifiable data.

Once the data were obtained, other challenges were identified, including different interpretations of identifiable characteristics between sites (e.g., BMI and stillbirth were deemed identifiable in some sites), long administrative delays, missing data and delays in addressing data queries, different definitions and capture of variables in the databases, and differences in diagnostic criteria used to diagnose GDM before and during the COVID pandemic.

Addressing social and economic differences through population databases

Where we are born and the family circumstances into which we are born are well documented to be associated with health outcomes. Dr Elpida Vounzoulaki, Dr Clare Gilles and colleagues from the University of Leicester discussed their experiences using primary care data through the Clinical Practice Research Datalink (CPRD) to demonstrate ethnic and socioeconomic differences in screening for T2DM after GDM. CPRD collects anonymised data from GP surgeries across the UK. Primary care data are linked to various other health-related databases providing a longitudinal population health dataset.

Challenges using these data for research include issues around data completeness, with missing data not missing at random, identifying who is not in the dataset and why, and the cost of accessing data, which can be over £100 000 for an institutional licence.

Lessons learned from creating a life-course, population-based research database in South London

An advantage of routinely collected data is that it can provide a more representative sample of the population, as data are collected from real-life health records rather than from a specific group of participants recruited for a study. Prof Lucilla Poston described her experience establishing the early life cross-linkage in research (eLIXIR) living cohort. This cohort combines data from several NHS sources (primary care, mental health, secondary care, social care) covering individuals living in several South London boroughs, representing a socioeconomically diverse, urban, multi-ethnic population [26]. There have been several challenges in bringing together this cohort, including administrative delays, obtaining funding from the Medical Research Council, creating a trusted research environment (TRE), obtaining ethical approvals and permissions, and allowing patients to opt out providing their data. A dedicated data entry and linkage person in the Trust has proven vital in addressing downstream issues with the data itself. This cohort has the potential to provide valuable insights into the health of this specific population and to evaluate interventions at the population level.

The potential for data to improve clinical risk prediction in GDM and other pregnancy conditions

Dr Lucy Mackillop discussed the development, translation and subsequent commercialisation of **GDm-Health**. Initially developed in Oxford, this system allows patients to digitally track their blood glucose levels and annotate the data with meal tags, medication doses, and comments. In addition, healthcare professionals can review submitted data and provide feedback via messages [27–29]. Over half of the National Health Service (NHS) trusts in England are now using the system. Data from the system have the potential for clinical decision support, with algorithms developed to predict the need for medication and risk of emergency caesarean birth [30,31]. The first algorithm, SYNE-GDM, predicts which women will likely require medication for their GDM in the next week. This product is UK Conformity Assessed (UKCA) marked but has yet to be deployed in clinical practice.

The Pre-eclampsia Integrated Estimate of RiSk (PIERS) is a model predicting adverse maternal outcomes, and was developed and validated in a clinical trial setting [32]. The model has been further refined using machine learning methods (PIERS-AI), with improved prediction of preeclampsia and other complications. Prof Laura Magee discussed the experience of conducting large-scale international research on pregnancy outcomes and the challenges of selecting meaningful outcomes for model development that are meaningful to pregnant women and communities where the application is intended.

Choosing the right statistical tool for the job

Despite machine learning approaches often being considered as extensions of traditional medical statistics methods, different disciplines focus on each method (i.e., biostatistics *vs* data science). These disciplines define terms differently, creating barriers to interdisciplinary learning. Dr. Lei Clifton discussed the advantages and challenges of working across methodologies and how this has led to breakthroughs in her work on polygenic risk scores for breast cancer. Dr Clare Gilles suggested that machine learning approaches could support biostatistical analyses by addressing the most significant challenge with many big datasets: missing data.

Examples of machine learning approaches that have been used in GDM were presented. Dr Durga Parkhi has studied continuous glucose monitoring data in women with GDM. Ms Jenny

Yang applied Logistic Regression and XGBoost Regression to predict the risk score of hyperglycaemia alerts for three days in women with GDM. In a validation dataset from a different hospital, she demonstrated her approach to be replicable. However, she also discussed challenges she experienced using the GDm-Health data, including missing data and variability in the duration that each woman used the system.

Identifying the appropriate research method or a “good” vs “bad” model can be difficult. There is a need for internal and external validation in different populations to minimise overfitting and evaluate the model’s performance in terms of its accuracy and generalisability. Training models on large datasets requires powerful machines for collection, storage, and analysis. The datasets are messy and large, requiring complex analysis and experienced analysts. When a model is identified and validated, turning it into a “product” for use in the health system, and then scaling it is not easy. Additionally, to clinically use any AI clinical decision tool, the algorithm must be approved as a Class 2B medical software device, which is expensive and time-consuming. Working together with industry partners is a key part of translation and implementation. Dr Mert Aral, from Huma, a British healthcare technology company that acquired GDm-Health in 2022, described their approach to working with researchers and clinicians to get AI advances into clinical use.

Key recommendations

Six recommendations emerged from the Innovation forum for future investment to enable research in GDM with relevance to many other areas of medicine (see Figure 1).

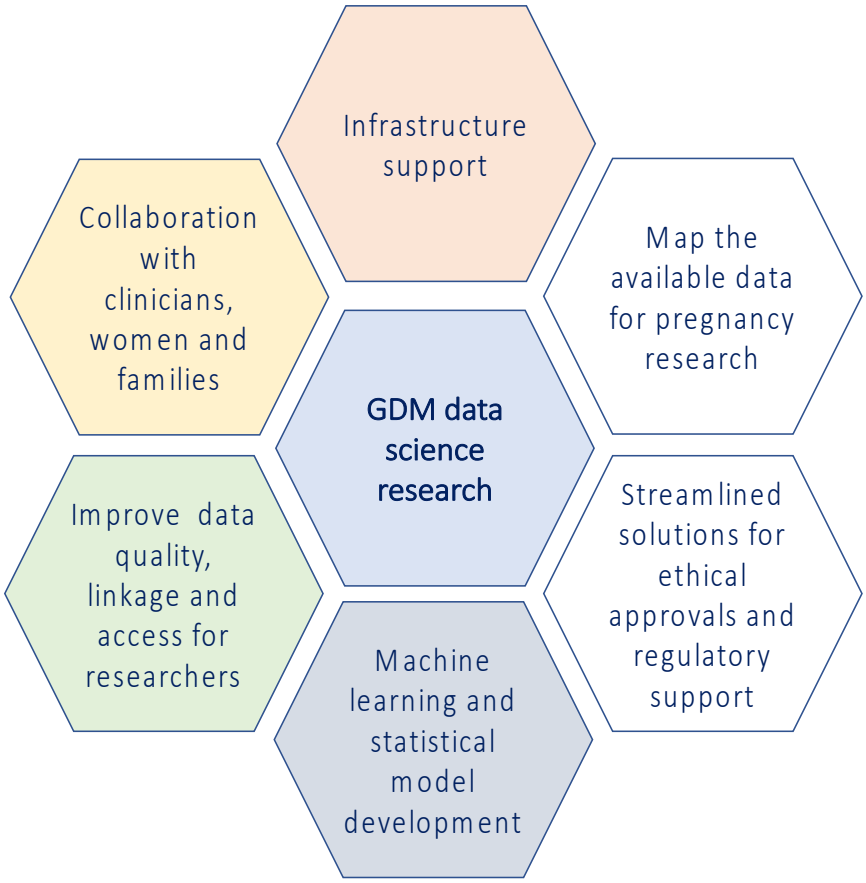

Figure 1: Key recommendations needed for data science research in GDM to improve clinical outcomes

Building the Infrastructure to enable complex clinical data research in GDM

Receiving approvals to access clinical data and then maintaining compliance is highly bureaucratic and time-consuming. Data management systems and sharing protocols must be secure, with a need for additional training for staff, such as on the use of trusted research environments (TREs), which are now known in England as secure research environments (SREs). TREs/SREs have gained popularity in the UK following the COVID-19 pandemic; however, they have limitations such as limited transparency, statistical software availability and difficulty sharing outputs.

De-identifying routine data can present challenges, such as identifiers stored in text format, making it impossible to check errors, and limiting future data linkage. Data confidentiality must be ensured when collecting, storing, and analysing data. Federated learning models, where algorithms developed centrally are trained on data stored locally without exchanging it to a central location, could hold future promise. However, given the heterogeneity in data systems and lack

of resources for data analytics for research in many centres, this may not be feasible within an NHS environment for maternity at this time.

Infrastructure investment and support are needed to enable the secure linkage of clinical datasets from maternity hospitals, NHS Digital, primary health care databases (such as CPRD) and third-party companies such as Huma. Other examples and potential routes to obtain this support were discussed.

Mapping routinely collected data available in the UK for pregnancy research.

Many routinely collected clinical data sources relevant to GDM research exist in the UK. However, it can be difficult for researchers to know which data are available, to access them and to obtain them in a format suitable for research. Whilst almost all NHS Trusts use electronic health record systems, these systems need to be interoperable. Data also resides in primary care databases, in local research platforms combining several data sources for a specific geographic or administrative area, national databases curated by NHS England equivalent devolved nations health boards, and third-party providers such as Huma, who own GDm-Health.

Building on the success of the national Neonatal Clinical Trials platform, Dr Ed Mullins and colleagues from Imperial College and the George Institute for Global Health are leading an initiative to establish a national maternity clinical trials platform. Hopefully, this platform will provide a user-friendly interface to link researchers with the appropriate anonymised data. Similar platforms in the US have facilitated large pragmatic clinical trials in GDM research [33]. It is hoped that in the UK, we can leverage routinely collected data to create a powerful platform to facilitate high-quality research on pregnancy in this country.

Streamlining the ethical approval process for studies involving data from many sites

As illustrated in the case study on GDM during the COVID-19 pandemic, the need for site-specific ethical approvals can significantly add to the administrative and paperwork burden of conducting multi-site studies. To move towards decentralised research methods such as federated learning, there must be agreement on data format and recording to enable local training of centrally developed models. This approach has been used extensively in other areas of data science; however, to our knowledge, it has not been used in the UK for pregnancy research.

Strategies to address data quality, linkage and access for researchers

Routine clinical data are not the same as those collected for research purposes. Frequently, key variables important for research may not be recorded, as they were not clinically relevant at the time of collection. The data can also include large numbers of missing, incomplete, wrong, inconsistent, or duplicate entries or data captured in free text. However, improving data collection could burden already overwhelmed NHS staff and services.

Much of the time spent on complex data analysis is taken up by data cleaning. The process to manipulate and clean the dataset is computationally intensive and time-consuming due to the poor data quality and the complexity of data structures and formats. Developing methods to streamline this process and learning from data scientists who have developed these approaches could save time and money. Luis Santos from the Alan Turing Institute discussed their approach to data integration, standards and security.

The linkage of datasets can be prohibitively costly. There may be much overlap between different datasets, such as when a patient changes GPs. There are differences in how data are collected and recorded across different organisations, making it difficult to compare and combine data from different sources. Each data source has a different approach for data capture and different methods for data extraction with varying levels of granularity. The lack of interoperability makes international collaboration difficult and risks contributing to global research inequities in this field.

Machine learning and statistical model development

Clinical machine learning algorithms and statistical models can be used to identify patterns (digital markers), patient subtypes, and quantify and map the associations among clinical confounders. Machine learning algorithms can overcome the limitation of handling high-dimensional and/or time-series data compared to statistical models; however statistical models have better explainability. Most clinical researchers are unaware of the potential and possibilities for machine learning and other statistical methods in addressing clinical challenges. The Forum highlighted the need to work together and across traditional siloed boundaries.

Machine learning has the potential to predict mothers with a high risk of developing GDM and monitor the status of mothers with GDM during and after pregnancy (e.g., blood glucose, medication, lifestyle and food intake). Tree-based methods, boosting-based methods, clustering methods and artificial neural network models have been used to predict blood glucose and assist clinicians in personalised patient management. Another approach is to use generative models such as generative adversarial networks (GANs) to represent patient information and thereby enable transfer learning – a type of machine learning model that can learn new knowledge by solving similar problems. This would allow the GDM community to build a knowledge map of GDM among hospitals nationwide.

Challenges persist in pushing the traditional siloed boundaries in two ways. First is the healthcare data infrastructure. Developing, tuning and applying machine learning models in clinical practice requires infrastructure to access the data. Partnering with data centres and industry partners will be essential for sustainability. Federated learning or blockchain may be solutions to the challenges in data privacy. The second challenge is the lack of ethical AI law and governance guidance for AI, which is a universal challenge for clinical machine learning applications. AI-based decision-making systems are classified as Type II medical devices. Clinical machine learning models need to be updated on-line (real-time response) or off-line (batch learning), which need to follow a model update protocol and decision rules. Clinical machine learning in GDM would require its own episode and criteria in machine learning decision risk/efficacy assessment.

Involving women and their families and clinicians in setting priorities for data science research in GDM

Whilst consumer groups were absent during the NCRM Innovation Forum, given the technical focus of discussions, it was acknowledged by all that to move this agenda forward, they must be central to future projects deciding what to investigate and interpreting the findings of the research. It will be essential to seek a range of opinions from women and their families representing different social and ethnic backgrounds and geographic regions across the UK if this research is to impact issues important to them.

Conclusions

Data science promises more effective, targeted, and affordable strategies for GDM research and clinical practice. The forum highlighted the need for individualized care pathways for women with GDM and the urgent need to improve the prediction and prevention of T2DM after birth. The recently commissioned National GDM Audit is expected to shine a spotlight on GDM care provision in the UK, with anticipated wider recognition of the importance of this condition during pregnancy and for women's lifelong health amongst the public, health workers, policymakers and research funders.

This forum focused on complex data for pregnancy and GDM research, and we highlight in this report several methodological challenges we believe have relevance beyond GDM. We also present learning from other disciplines, emphasising the importance of working across traditional boundaries and with industry partners. A collective voice, including service users, is needed if we are to gain investment in research and data-linkage infrastructure for GDM in the UK and make a step change in care experiences and outcomes for all women.

Innovation Forum participants

Name	Affiliation
Jane Hirst	University of Oxford
Lucy Mackillop	OUH NHS Foundation Trust/University of Oxford/EMIS Health
Huiqi Yvonne Lu	University of Oxford
Ponnusamy Saravanan	University of Warwick
Sara White	Kings College London
Yasmina Al Ghadban	University of Oxford
Mert Aral	Huma
Nerys Astbury	University of Oxford
Lei Clifton	University of Oxford
Helen Duncan	Department of Health and Social Care
Clare Gilles	University of Leicester
Trudie Lang	The Global Health Network/University of Oxford
Nick Lewis-Barned	Northumbria Healthcare NHS Foundation Trust (retired)/ National Diabetes Audit
Ping Lu	University of Oxford
Laura Magee	Kings College London
Anastasios Malakasis	OUH NHS Foundation Trust
Bilal Mateen	Wellcome Trust
Alaffia Maxwell	Huma
Ed Mullins	Imperial College London
Durga Parkhi	University of Warwick
Lucilla Poston	Kings College London
Rebecca Reynolds	University of Edinburgh
Jenny Yang	University of Oxford
Elpida Vounzoulaki	University of Leicester
Tingting Zhu	University of Oxford
Luis Santos	Alan Turing Institute
Eleanor Scott	University of Leeds
Katherine Ripullone	OUH NHS Foundation Trust

References

1. Wang H, Li N, Chivese T, Werfalli M, Sun H, Yuen L, et al. IDF Diabetes Atlas: Estimation of Global and Regional Gestational Diabetes Mellitus Prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's Criteria. Diabetes Research and Clinical Practice. 2022 Jan 1;183.

2. Ye W, Luo C, Huang J, Li C, Liu Z, Liu F. Gestational diabetes mellitus and adverse pregnancy outcomes: systematic review and meta-analysis. BMJ. 2022;377.

3. Lauenborg J, Hansen T, Jensen DM, Vestergaard H, Mølsted-Pedersen L, Hornnes P, et al. Increasing incidence of diabetes after gestational diabetes: a long-term follow-up in a Danish population. Diabetes Care. 2004 May;27(5):1194–9.

4. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet. 2009 May 23;373(9677):1773–9.

5. Daly B, Toulis KA, Thomas N, Gokhale K, Martin J, Webber J, et al. Increased risk of ischemic heart disease, hypertension, and type 2 diabetes in women with previous gestational diabetes mellitus, a target group in general practice for preventive interventions: a population-based cohort study. PLoS medicine. 2018;15(1):e1002488.

6. Leybovitz-Haleluya N, Wainstock T, Landau D, Sheiner E. Maternal gestational diabetes mellitus and the risk of subsequent pediatric cardiovascular diseases of the offspring: a population-based cohort study with up to 18 years of follow up. Acta Diabetol. 2018 Oct;55(10):1037–42.

7. Kaseva N, Vääräsmäki M, Matinolli HM, Sipola-Leppänen M, Tikanmäki M, Heinonen K, et al. Pre-pregnancy overweight or obesity and gestational diabetes as predictors of body composition in offspring twenty years later: evidence from two birth cohort studies. Int J Obes (Lond). 2018 Apr;42(4):872–9.

8. Nijs H, Benhalima K. Gestational diabetes mellitus and the long-term risk for glucose intolerance and overweight in the offspring: a narrative review. Journal of clinical medicine. 2020;9(2):599.

9. Zhao P, Liu E, Qiao Y, Katzmarzyk PT, Chaput JP, Fogelholm M, et al. Maternal gestational diabetes and childhood obesity at age 9-11: results of a multinational study. Diabetologia. 2016 Nov;59(11):2339–48.

10. Yu Y, Arah OA, Liew Z, Cnattingius S, Olsen J, Sørensen HT, et al. Maternal diabetes during pregnancy and early onset of cardiovascular disease in offspring: population based cohort study with 40 years of follow-up. BMJ. 2019 Dec 4;367:l6398.

11. Clausen TD, Mathiesen ER, Hansen T, Pedersen O, Jensen DM, Lauenborg J, et al. Overweight and the Metabolic Syndrome in Adult Offspring of Women with Diet-Treated Gestational Diabetes Mellitus or Type 1 Diabetes. The Journal of Clinical Endocrinology & Metabolism. 2009 Jul 1;94(7):2464–70.

12. Chen S, Zhao S, Dalman C, Karlsson H, Gardner R. Association of maternal diabetes with neurodevelopmental disorders: autism spectrum disorders, attention-deficit/hyperactivity

disorder and intellectual disability. International Journal of Epidemiology. 2021 Apr 1;50(2):459–74.

13. Nomura Y, Marks DJ, Grossman B, Yoon M, Loudon H, Stone J, et al. Exposure to gestational diabetes mellitus and low socioeconomic status: effects on neurocognitive development and risk of attention-deficit/hyperactivity disorder in offspring. Arch Pediatr Adolesc Med. 2012 Apr;166(4):337–43.

14. Kong L, Nilsson IAK, Brismar K, Gissler M, Lavebratt C. Associations of Different Types of Maternal Diabetes and Body Mass Index With Offspring Psychiatric Disorders. JAMA Netw Open. 2020 Feb 5;3(2):e1920787.

15. Saravanan P, Magee LA, Banerjee A, Coleman MA, Von Dadelszen P, Denison F, et al. Gestational diabetes: opportunities for improving maternal and child health. The Lancet Diabetes & Endocrinology. 2020 Sep 1;8(9):793–800.

16. Recommendations | Diabetes in pregnancy: management from preconception to the postnatal period | Guidance | NICE [Internet]. NICE; [cited 2022 Oct 28]. Available from: https://www.nice.org.uk/guidance/ng3/chapter/Recommendations

17. Chivese T, Hoegfeldt CA, Werfalli M, Yuen L, Sun H, Karuranga S, et al. IDF Diabetes Atlas: The prevalence of pre-existing diabetes in pregnancy – A systematic reviewand meta-analysis of studies published during 2010–2020. Diabetes Research and Clinical Practice. 2022 Jan 1;183:109049.

18. Lawrence JM, Contreras R, Chen W, Sacks DA. Trends in the Prevalence of Preexisting Diabetes and Gestational Diabetes Mellitus Among a Racially/Ethnically Diverse Population of Pregnant Women, 1999–2005. Diabetes Care. 2008 May 1;31(5):899–904.

19. Stacey T, Tennant P, McCowan L, Mitchell E, Budd J, Li M, et al. Gestational diabetes and the risk of late stillbirth: a case–control study from England, UK. BJOG: An International Journal of Obstetrics & Gynaecology. 2019;126(8):973–82.

20. Sovio U, Murphy HR, Smith GCS. Accelerated Fetal Growth Prior to Diagnosis of Gestational Diabetes Mellitus: A Prospective Cohort Study of Nulliparous Women. Diabetes Care. 2016 Apr 7;39(6):982–7.

21. Venkataraman H, Ram U, Craik S, Arungunasekaran A, Seshadri S, Saravanan P. Increased fetal adiposity prior to diagnosis of gestational diabetes in South Asians: more evidence for the ‘thin–fat’ baby. Diabetologia. 2017 Mar 1;60(3):399–405.

22. Farrar D, Fairley L, Santorelli G, Tuffnell D, Sheldon TA, Wright J, et al. Association between hyperglycaemia and adverse perinatal outcomes in south Asian and white British women: analysis of data from the Born in Bradford cohort. Lancet Diabetes Endocrinol. 2015 Oct;3(10):795–804.

23. Ayman G, Strachan JA, McLennan N, Malouf R, Lowe-Zinola J, Magdi F, et al. The top 10 research priorities in diabetes and pregnancy according to women, support networks and healthcare professionals. Diabet Med. 2021 Aug;38(8):e14588.

24. Beck RW, Bergenstal RM, Laffel LM, Pickup JC. Advances in technology for management of type 1 diabetes. The Lancet. 2019 Oct 5;394(10205):1265–73.

25. Thagaratinam S, Khan R, Blott M, Nelson-Piercy C, Williams D, Relph S, et al. Guidance for maternal medicine in the evolving coronavirus (COVID-19) pandemic.

https://www.rcog.org.uk/media/nkpfvim5/2020-12-09-guidance-for-maternal-medicine-services-in-the-coronavirus-c.pdf: Royal College of Obstetricians and Gynaecologists; 2020.

26. About – The eLIXIR Programme [Internet]. [cited 2023 Jan 9]. Available from: http://www.guysandstthomasbrc.nihr.ac.uk/microsites/elixir/about-the-programme/

27. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. J Diabetes Sci Technol. 2014 Nov 1;8(6):1105–14.

28. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the Efficacy of a Mobile Phone-Based Blood Glucose Management System With Standard Clinic Care in Women With Gestational Diabetes: Randomized Controlled Trial. JMIR mHealth and uHealth. 2018 Mar 20;6(3):e9512.

29. Hirst JE, Mackillop L, Loerup L, Kevat DA, Bartlett K, Gibson O, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol. 2015 Jan;9(1):111–5.

30. Yang J, Clifton D, Hirst JE, Kavvoura FK, Farah G, Mackillop L, et al. Machine Learning-Based Risk Stratification for Gestational Diabetes Management. Sensors (Basel). 2022 Jun 25;22(13):4805.

31. Lu H, Hirst J, Yang J, Mackillop L, Clifton D. Standardising the assessment of caesarean birth using an oxford caesarean prediction score for mothers with gestational diabetes. Healthcare Technology Letters. 2022 Mar 13;9.

32. Dadelszen P von, Payne B, Li J, Ansermino JM, Pipkin FB, Côté AM, et al. Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model. The Lancet. 2011 Jan 15;377(9761):219–27.

33. Hillier TA, Pedula KL, Ogasawara KK, Vesco KK, Oshiro CES, Lubarsky SL, et al. A Pragmatic Randomized Clinical Trial of Gestational Diabetes Screening. N Engl J Med. 2021 Mar 11;384(10):895–904.

Appendix: Innovation Forum Program

NCRM Innovation Forum on Complex Clinical Data and GDM

Thursday 8th and Friday 9th December 2022

Ship Street Centre, Jesus College, Oxford

AGENDA

Thursday 8th December		
10.00	Arrival and registration	Rebecca Chaplin (University of Oxford)

Session 1: Introduction		
10.30 - 10.45	Welcome, introductions, objectives of the meeting	Jane Hirst (University of Oxford)
10.45 - 11.00	Introduction to the clinical problems managing women who develop gestational diabetes mellitus (GDM)	Ponusammy Saravanan (University of Warwick)

Session 2: Routinely collected data in pregnancy research: Challenges and successes		
11.00 - 11.15	The GDM- Health experience	Lucy Mackillop (Oxford University Hospitals NHS Foundation Trust)
11.15 - 11.30	COVID and GDM: challenges in using routinely collected data in GDM research in the UK	Rebecca Reynolds (University of Edinburgh)
11.30 - 11.45	Challenges measuring the longer-term effects of gestational diabetes in the UK	Elpida Vounzoulaki (University of Leicester)
11.45 - 12.00	Data lessons from preeclampsia research in the UK and abroad	Laura Magee (King's College London)
12.00 - 12.30	Panel discussion and questions	Moderator: Jane Hirst
12.30 – 13.30	Networking lunch	

Session 3: Challenges accessing NHS and other data sources for research on women in pregnancy		
13.30 - 13.45	Challenges in a South East London maternal/child electronic health record linkage (eLIXIR), and the potential for using in clinical research.	Lucilla Poston (KCL)
13.45 - 14.00	Challenges with data used in primary care for research	Neil Martin (EMIS)
14.00 - 14.15	Lessons from the National Diabetes in Pregnancy Audit and National GDM audit pilot	Eleanor Scott (University of Leeds)
14.15 - 14.30	National maternity data challenges and opportunities for research collaborations	Helen Duncan (National Lead for Lifecourse Intelligence, Office for Health Improvement and Disparities, Department of Health and Social Care)
14.30 – 14.40	Use of routine data to answer clinical questions – A Funder's Perspective	Bilal Mateen (Wellcome Trust)
14.40 - 15.00	Panel discussion and questions	Moderator: Sara White

		(King's College London)
15.00 – 15.15	Afternoon coffee break	

Session 4: Challenges combining different data platforms for research and clinical purposes		
15.15 - 15.30	Machine Learning Approaches for Clinical Timeseries Data	Tingting Zhu (RAEng research Fellow, University of Oxford)
15.30 - 15.45	Establishing a National Maternity research data platform	Ed Mullins (Imperial College)
15.45 - 16.00	Linking electronic health records across Manchester: challenges and lessons for research	Jenny Myers (University of Manchester)
16.00 - 16.15	Challenges integrating personal and third-party data into EHR for clinical and research use	Aral Mert (Huma)
16.15 - 16.30	Using clinical data in Global Health and the Global Health Network	Trudie Lang (University of Oxford)
16.30 - 17.00	Panel discussion and questions	Moderator L Mackillop
17.00 - 17.30	Day 1 wrap up and discussion about a joint publication to come from meeting	
17.30 - 18.30	Drinks reception - CB Tower Room, Cheng Yu Tung Building	

Friday 9th December

Session 5: Methodologic and analytic challenges: machine learning &classic statistical modelling		
8.45 - 9.00	Morning coffee and refreshments	
9.00 - 9.20	Challenges combining machine learning with medical statistics	Lei Clifton (University of Oxford
9.20 - 9.40	Statistical challenges in diabetes research in the UK	Clare Gilles (University of Leicester)
9.40 - 9.55	ML approaches for Continuous glucose data in pregnancy	Durga Parkhi (University of Warwick)
9.55 - 10.10	ML approaches for intermittent glucose data	Jenny Yang (University of Oxford)
10.10 – 10.25	Challenges in data-centric engineering for AI in healthcare	Ann-Marie Mallon (The Turing Institute)
10.25 - 10.50	Panel discussion and questions	Moderator: Yvonne Lu (University of Oxford)
10.50 - 11.00	Morning coffee break	

Session 6: Bringing it together		
11.00 – 12.00	What are the key gaps in data access, ability to combine different data sources, methodologic challenges? **Where to next?**	Moderator: Jane Hirst All to contribute
12.00	Meeting close	