AIMS For a better birth...

Women, Pregnancy and Artificial Intelligence: Opportunities and Cautions in the Age of Digital Maternity

AIMS Journal, 2025, Vol 37, No 4

By Christopher Yau, Nuffield Department for Women's & Reproductive Health, University of Oxford on behalf of the MUM-PREDICT and OPTIMAL Consortiums

The MUM-PREDICT[1] consortium is a team of experts in Data Science, Obstetrics, Psychological Medicine, Primary Care, and Public Health working alongside women with experience of having two or more health conditions that presented before pregnancy. Supported by the Medical Research Council (MRC), our consortium sought to conduct research to improve care for pregnant women with two or more long-term health conditions. In parallel, a number of the team members were also supported by the National Institute for Health Research (NIHR) to conduct research into how artificial intelligence - or AI - could be used to improve our understanding of multiple long-term conditions in the general population for a project called OPTIMAL. In this article, I will discuss the context and background to the two projects and the AI technologies that have emerged from our research.

What is artificial intelligence?

AI has rapidly entered into our daily lives and discussions of the implications of new AI technologies are now in the mainstream of public thought. At its simplest, AI refers to computer systems that are able to mimic human reasoning and creativity. These could be chatbots that respond to questions in a natural, human-like way. It could be software that can produce photorealistic images and videos, based on

nothing more than a simple command such as “create a photo of a family on a beach”. More recently, it has become possible to set up AI “agents” to act as personal assistants to manage our diaries, book transport or even order flowers to be delivered for loved ones on special occasions. AI systems are also working behind the scenes to, for example, scan financial transactions to identify potential fraudulent activity. These new uses of AI have arisen due to a cluster of advances within the last 10-15 years. These include the development of bigger and more powerful computers that can support more substantial computer models that contain and process vast amounts of information and knowledge.

It is therefore unsurprising that the health system has also taken a keen interest in AI. Advocates suggest that AI can offer solutions to support earlier detection of health conditions, reduce workloads on staff, and help to drive efficiencies and standardise care. For example, using AI to read and interpret ultrasound scans and fetal heart rate monitoring during labour could produce more consistent, “objective”[2] readings, reducing variability amongst different human operators, minimising human error, or to flag potential problems more quickly. However, there are still relatively few AI tools in widespread use within the National Health Service due to the lack of long-term evidence for the efficacy of these AI tools, lack of computing facilities and regulatory barriers.

Artificial intelligence for studying Multiple Long-Term Conditions

One area where AI is gaining traction is in predicting health risks. AI algorithms can now scan through thousands of health records, test results, and other data to estimate who is more likely to develop conditions such as pre-eclampsia, gestational diabetes, or experience pre-term labour. These “risk prediction models” can support clinicians to provide more tailored care and give expectant mothers a better understanding of their own personal risks rather than an average risk which is applicable over the wider population. This is particularly important when we consider women entering pregnancy with multiple long-term conditions (MLTCs).

More than ever, maternity services are caring for women with layered health needs that do not fit neatly into standard pathways. Around one in five pregnant women in the UK now lives with two or more long-term health conditions—whether physical, such as diabetes or asthma, or mental, such as anxiety or depression. These women may be more likely to encounter complications during pregnancy, birth, or the postnatal period[3] (Lee et. al. (2023), but standard risk prediction models, and traditional care pathways, often struggle to capture the nuance of their needs. For example, pregnancy-related complications and reproductive factors are associated with an increased risk of cardiovascular disease (CVD) later in life. The QRISK®-3 is a well established risk prediction tool that can be used to identify individuals at a higher risk of developing cardiovascular disease. It has been well tested and validated on the general population. However, QRISK®-3 does not make use of pregnancy-related factors (e.g. gestational hypertension, preeclampsia, miscarriage) which might change a women’s individual risk for CVD. In MUM-PREDICT, we created a modification of the QRISK-3 model to include pregnancy factors which gave improved predictions.[4]

Research on maternal health during pregnancy has historically focused on individual diseases rather than

the co-existence of multiple long-term conditions. In MUM-PREDICT, we examined anonymised general practice (GP) data for 242,678 pregnant women with multiple long-term conditions and found 33,646 unique combinations of health conditions amongst this group. Some combinations were much more common than others but there existed a wide variety of patterns of health conditions. In order to further study so many patterns, it was necessary to develop our own AI algorithms to summarise and determine which of the 33,646 combinations of MLTCs were most important or representative (Figure 1).

Our AI algorithm - called "mmVAE"[5] - enabled us to find sixty-six "clusters" - groups of individuals who have similar combinations of two or more health conditions - which were present in at least 1% of the women studied. Eight of these clusters involved different combinations of the seven of the most prevalent health conditions found in pregnant women with MLTCs (depression, anxiety, allergic rhinoconjunctivitis, asthma, migraine, irritable bowel syndrome and mental health conditions). Some involved well-known conditions, such as endometriosis, while others involved combinations of rare health conditions, whose existence could only be identified through assembling a large number of pregnancy related data. Our work also showed that different clusters were associated with different risks of miscarriage, and understanding these patterns could support clinicians to devise more specific approaches to the care offered when presented with complex presentations.

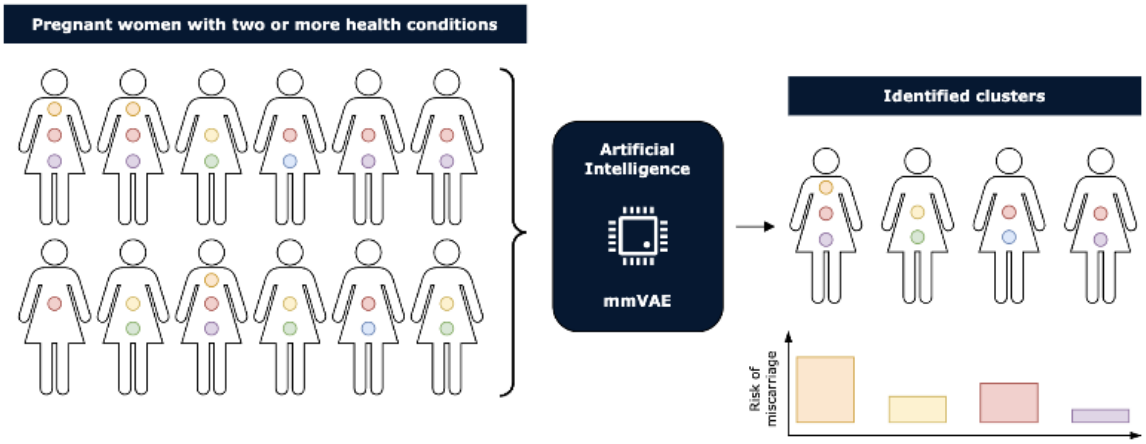

Figure 1: Clustering women with two or more health conditions during pregnancy. Artificial intelligence can be used to identify groups of women with similar health conditions in order to assess their group's risk of complications such as miscarriage.

Chatting with patient data

Scanning health records using AI to find useful patterns can therefore be highly informative, but it is not very interactive for clinicians or patients. Recent advances in AI for language processing have given rise to Large Language Models (or LLMs). These offer a more accessible gateway for healthcare professionals and the public alike to access health information. For example, some healthcare providers are experimenting with AI-driven chatbots and digital assistants to support pregnant women outside of appointments. These tools can answer basic questions, give reminders, or offer symptom checkers. For some women - especially those with limited access to care - they may be particularly helpful.

LLMs are trained to analyse sequences of words, identify patterns and predict what the next words should be. By learning from massive quantities of text (usually from the internet), LLMs can identify the rules of language such as syntax and grammar, and apply these to understand and also form long passages of realistic text. This has given rise to software, like ChatGPT, that can respond in human-like ways to natural language questioning including medical issues, since these AI applications have extensively scanned the published medical literature. Recent surveys suggest many GPs are already using LLMs to support their work.[6]

Despite their increasingly widespread use, most Large Language Models are actually not approved for medical use or for providing advice about medical matters. For example, an LLM might provide guidance on whether it is safe to take certain medications during pregnancy, but this advice may not reflect individual risks and should never replace consultation with a qualified healthcare professional. Some LLMs actively resist responding to medical queries, while others provide qualified information, stating that they are sharing evidence-based information and providing an interpretation of published population statistics.

The reason why caution must be applied when using Large Language Models for medical purposes is that they have a number of limitations. The responses they generate can contain “hallucinations”. This is information that an LLM has synthesised from what it has learnt, but is actually factually incorrect. For instance, an LLM might incorrectly claim that a common over-the-counter medication is completely safe in all trimesters of pregnancy, when in reality it may pose risks during the first trimester. Responses can also be biased. LLMs may be skewed toward certain perspectives or ideas because they are more prominent in the literature they have been trained on, rather than reflecting actual consensus or individual variability. For example, advice about dietary supplements or natural remedies during pregnancy might disproportionately reflect sources that promote alternative medicine, even if the majority of clinical guidelines do not support their use. This makes LLMs vulnerable to large volumes of medically-unverified information available on the internet.[7]

When answering complex questions, the reasoning capabilities of LLMs can break down. An LLM may misunderstand a question, or the chain of thought may become confused. In a recent study, researchers found that ChatGPT could provide the correct response to questions such as “what are the symptoms of pre-eclampsia” approximately 70% of the time. However, when the questions required more reasoning, such as deciding on the best course of action for a pregnant woman showing signs of placental abruption,

its performance significantly declined, with only 50% of the questions answered correctly.[8]

Keeping data distant

Nonetheless, the potential of LLMs is enormous and considerable research is taking place to make LLMs safer for health-related applications. However, the developers of LLMs can be limited in what they can accomplish. Published medical literature only contains summaries of the key findings from scientific and clinical studies. They do not contain the original individual-level patient data that may have been used to obtain those study results. This limits the detail with which LLMs can report risks. This may be a good thing. LLMs can be very powerful and have the capacity to absorb entire databases of health records - this means they could "memorise" patient data. Once memorised, the possibility exists that the entirety of that record could then be revealed, if the correct question is asked of it, and the individual's privacy violated, even if steps to anonymise the information have been taken. For this reason, many health data providers are putting in place explicit instructions to disallow the use of LLMs for processing of individual-level patient data.

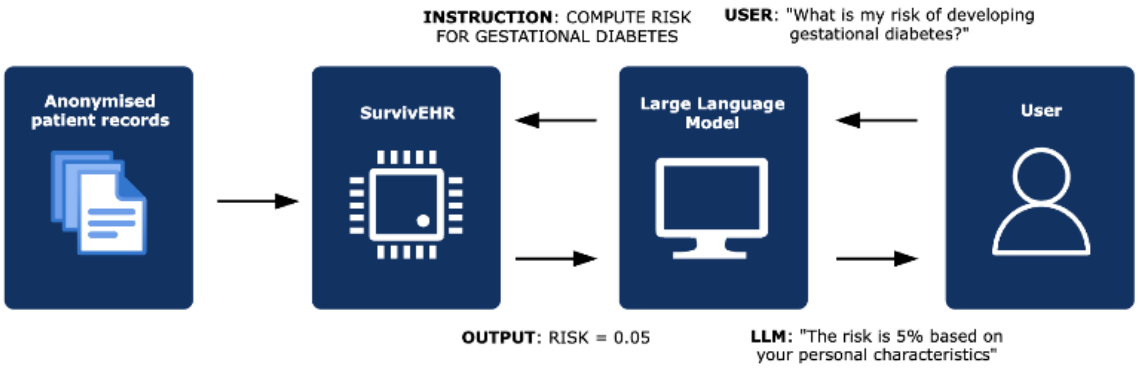

Figure 2: Separating data from Large Language Models with SurvivEHR. Large Language Models can interact with SurvivEHR to obtain answers to health risk questions posed by a user. SurvivEHR uses patient record information to provide risk levels. The Large Language Model does not interact with the patient records directly.

However, rather than allowing LLMs direct access to patient records, we could instead provide an intermediary - a computer program that is able to see the patient records and can provide LLMs with specific person-specific health risk information. Within the OPTIMAL consortium we developed "SurvivEHR" (pronounced 'survivor'), an AI tool that allows us to compute the risk of individuals acquiring up to 80 different health conditions, by learning from patterns it has seen in 23 million patients and their anonymised GP records. In future, Large Language Models could interface with SurvivEHR to get accurate patient-derived risk predictions based on solid evidence, but without being directly exposed to the health records and the privacy and data security concerns that arise (Figure 2).

What next?

The pace of AI advancement has been astonishing and the work of the MUM-PREDICT and OPTIMAL

consortiums represents just the beginning of what AI can accomplish in maternal healthcare. Moving forward, several key priorities must guide our efforts: rigorous validation of AI tools in diverse populations, ongoing monitoring for bias and fairness, robust privacy protection measures, and meaningful involvement of women and families in the development process. Healthcare providers, researchers, and policymakers must work together to create frameworks that harness AI's potential while safeguarding against its risks. Most importantly, we must ensure that the voices and experiences of women - particularly those with multiple long-term conditions - remain at the centre of these technological advances. Only through this collaborative approach can we build an AI-enhanced healthcare system that truly serves the needs of all women during one of the most important periods of their lives.

Finally, if you are reading this you might be wondering what all this means for you right now. The short answer is: not much has changed yet, but there is reason to be both hopeful and cautious about what is coming. Many AI tools that can be specifically used for medical purposes are still being developed and tested and most people will not encounter them just yet in a routine healthcare setting. However, when these medical AI technologies do become widely available, we can expect that they will make care more tailored to specific situations. Instead of hearing about risks based on all pregnant women, you might get information that considers your particular combination of conditions - whether that's asthma and migraines, or diabetes and anxiety. In the interim, there is no harm in safely experimenting with AI technologies but just be aware of its limitations and check with an appropriate healthcare professional before taking any actions.

Author Bio: Christopher Yau is Professor of Artificial Intelligence based at the Big Data Institute in Oxford working across the Nuffield Department of Women's and Reproductive Health and the Nuffield Department of Population Health.

---

1 MUM-PREDICT Project. (2024). Multi-morbidity in pregnancy: understanding health trajectories and care pathways using data across the UK. https://www.mumpredict.org/

2 AI used to read and interpret ultrasound scans and fetal heart rate monitoring during labour, could produce more consistent readings, reducing variability amongst different human operators, minimising human error and flagging up potential problems more quickly. AI can be considered to be more objective as a result of these behaviours - BUT - there remains issues around potential biases in the data and what values are used to develop the AI and therefore its objectivity is questionable.

3 Lee, S.I., Hope, H., O'Reilly, D., Kent, L., Santorelli, G., Subramanian, A., Moss, N., Azcoaga-Lorenzo, A., Fagbamigbe, A.F., Nelson-Piercy, C. and Yau, C., 2023. Maternal and child outcomes for pregnant women with pre-existing multiple long-term conditions: protocol for an observational study in the UK. BMJ open, 13(2), p.e068718. https://www.researchgate.net/publication/368808420_Maternal_and_child_outcomes_for_pregnant_wo

men with pre-existing multiple long-term conditions protocol for an observational study in the UK

4 Wambua, S., Crowe, F.L., Thangaratinam, S., O'Reilly, D., McCowan, C., Brophy, S., Yau, C., Nirantharakumar, K., Riley, R.D., Snell, K.I. and MuM-PreDiCT Group, 2025. Development and validation of a postpartum cardiovascular disease risk prediction model in women incorporating reproductive and pregnancy-related predictors. *BMC medicine*, *23*(1), p.508. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-025-04229-1

5 Gadd, C., Nirantharakumar, K. and Yau, C., 2022, November. mmVAE: multimorbidity clustering using Relaxed Bernoulli $ $-Variational Autoencoders. In Machine Learning for Health (pp. 88-102). PMLR. https://share.google/BoM0W0iNfuJz0rlck

6 Blease, C.R., Locher, C., Gaab, J., Hägglund, M. and Mandl, K.D., 2024. Generative artificial intelligence in primary care: an online survey of UK general practitioners. BMJ Health & Care Informatics, 31(1), p.e101102.

https://pmc.ncbi.nlm.nih.gov/articles/PMC11429366/

7 Han, T., Nebelung, S., Khader, F., Wang, T., Müller-Franzes, G., Kuhl, C., Försch, S., Kleesiek, J., Haarburger, C., Bressem, K.K. and Kather, J.N., 2024. Medical large language models are susceptible to targeted misinformation attacks. NPJ digital medicine, 7(1), p.288. https://www.nature.com/articles/s41746-024-01282-7

8 Bachmann, M., Duta, I., Mazey, E., Cooke, W., Vatish, M. and Davis Jones, G., 2024. Exploring the capabilities of ChatGPT in women's health: obstetrics and gynaecology.*npj Women's Health*, *2*(1), p.26. https://www.nature.com/articles/s44294-024-00028-w