MathSys miniproject: Identifying epidemiological causal factors of recurrent pregnancy loss (RPL)
Supervisor: Burroughs (Mathematics). Second supervisor: Quenby (Medical School) and/or Martyn Plummer (Statistics).

Background. Recurrent pregnancy loss (RPL) is a poorly understood health condition that affects 15% of women, and is characterized by repeated miscarriage. Although prognosis for final successful preganancy is fair, with around 65% of women with RPL acheiving a live birth, this may involve large numbers of previous miscarriages (10 or more). Hence, understanding the causal factors and modulating their impact would have a huge impact on these women. Currently there are a number of associations that have been identified (whether they are causal has not been ascertained). This includes increased RPL risk with body mass index (BMI) and smoking, whilst the probability of another miscarriage increases with the number of previous miscarriages, suggesting the condition is heterogeneous (degrees of severity).

Warwick (Quenby) is leading a massive RPL epidemiology research programme where couples register in the programme and are monitored through time. Extensive medical and demographic variables are measured, both initially and throughout the course of being in the programe. Currently there are 3000 registered patients. This is the largest and most extensive longitudinal study world-wide so provides a massive untapped resource.

The project. The project will involve two phases. Firstly, you will build a pregnancy/miscarriage events model incorporating potential causal factors. This can be similar to the preterm-birth model [1], but a stratification model capturing the severity heterogeneity is conceptually more realistic, where patients are in an (unknown and to be inferred) miscarriage severity class. The latter would explain the dependence on previous miscarriages; as more miscarriages occur, the evidence (probability) for that patient to be in a more severe class increases. Miscarriage is the combined probability of the probability of pregnancy and the probability of miscarriage given they are pregnant. RPL patients can in fact have either a low or high pregnancy probability further increasing the complexity of the problem. Analysis of these models would allow the causes and the impact of heterogeneity to be assessed in the condition.

Secondly, you will use Bayesian computational techniques (Markov chain Monte Carlo methods) to infer the model parameters from the data and identify the significant factors that determine the risk of RPL. This could be in a package, eg STAN, utilising suitable R packages, or developing bespoke code (eg in MatLab). If time permits you would use Gibbs variable selection to determine significant factors concommitently with parameter inference.

Desirable skills. An understanding of Markov processes and Markov chain Monte Carlo (MCMC) techniques is required. An ability to program in a high level language such as R or MatLab is essential. Other languages can be used if preferred, e.g. C++, python.

Opportunities for a PhD. This mini-project can lead to a PhD with Burroughs, Plummer and Quenby with UHCW as the external partner. There are multiple avenues for developing this into a PhD depending on interests:

- The PhD could develop the models and analysis, eg incorporate temporal observables such as illnesses, ultrasound scans and sequencing data of the miscarried fotuses. Miscarriage can result either from problems with the embryo (determined from sequencing) or with the uterus.

- A subset of the data also has biopsy data available which can be used to infer the state of the uterus for implantation. The PhD could focus on building and inferring a combined model of the epidemiology and cellular state.

- A healthcare focused PhD, where model based predictive tools would be developed to determine the impact of treatments and life-style changes, and thus determine optimal treatment. Such tools could be made available to the general public and clinicians. Linking to clinical trials may also be possible (collaboration with Birmingham).

References:
[1] Modelling the type and timing of consecutive events: application to predicting preterm birth in repeated pregnancies. Joanna H. Shih, Paul S. Albert. *Appl. Statist.* (2015), 64, Part 5, pp. 711730.