3) Computational disease progression models

1. Build computational models that combine multiple imaging data sources to describe phenotypes related to a hypertensive pregnancy in women and young adults.
2. Build computational models to describe hypertensive disease progression in UK Biobank imaging data including sub-trajectories of hypertensive disease.
3. Compare disease progression related to hypertensive pregnancy in women and young adults to patterns of hypertensive disease progression defined in UK Biobank.

**Workstream 3 (Lead: Lapidaire): Use computational models to study hypertensive disease progression in women and young adults after hypertensive pregnancy and how this relates to generalised hypertensive disease progression.**

**Workstream 3.1: *Develop computational model of disease progression in women and young adults after hypertensive pregnancy -*** As the completed dataset becomes available from Workstreams 1 we will start to develop computational models based on the cTI approach (see scientific importance section) to describe variation in imaging-derived phenotypes between women who have had a hypertensive pregnancy and those who have had only normotensive pregnancies. All women within the model will be positioned relative to variation within the most highly weighted features for the model. We will use the summary weighted scoring of individuals to explore the relative change in individual parameters between ‘health’ and ‘disease’ (disease progression) as well as how scores vary over time from index pregnancy. A separate model will be created from the South African dataset to investigate whether different disease patterns are evident. We will also apply the cTI algorithm to the combined datasets once available from Workstream 2. Within this dataset we will contrast young adults born to a hypertensive pregnancy to those born to a normotensive pregnancy. To explore the pattern of hypertensive disease progression in women and young adults we will quantitatively study relative weightings of biomarkers in the two models.

***Workstream 3.2 Develop computational model of hypertensive disease progression in UK Biobank -*** While data collection is in progress in Workstream 1 and 2 we will commence creation of computational models to describe the phenotypic landscape of hypertensive disease progression for all women in the general UK Biobank population, and compare this to landscapes in the men. Modelling will be performed with participants who have already had a heart attack or stroke excluded to ensure models represent multi-organ changes prior to acute events. Our current approach based on the initial 2,200 imaging datasets [Figure 2] has shown proof-of-principle for characterisation of hypertensive disease progression based on contrasts between participants with blood pressure >160/100mmHg and <120/80mmHg, including ability to generate a relative scoring that differentiates stages of disease progression for individual features. Our first objective will therefore be to replicate this work in the larger dataset of 20,573 female and in a dataset with 19,005 male participants we now have available. This larger sample size will also allow for identification of sub-trajectories in hypertensive disease progression. Over the course of the programme grant, the data sample is due to increase to 100,000 participants [48], with additional repeat imaging data, and linkage to hospital 10 event statistics and primary care records planned. Although we do not expect to need this scale of sample size our strategy will be replicate the models, and retrain if required, as new batches of data are released. In addition, as sample size increases, we expect to be able to more robustly define the key sub-trajectories in hypertensive disease development.

***Workstream 3.3: Place mothers in UK Biobank hypertensive disease model sub-trajectories-***As our study cohorts have undergone comparable study investigations as UK Biobank (Table 3), we will also place them in the phenotypic landscapes developed from the UK Biobank data. Working with our co-investigators in Montreal, we have developed a strategy for data standardisation and imputation within the cTI approach that enables us to add participants to the model without affecting parameter weightings. This allows us to generate positions of individual, new participants within the disease progression model. From this analysis we can assess where women following a hypertensive pregnancy, including subgroups of women from different backgrounds such as those from South Africa, and young adults born to a hypertensive pregnancy fall within disease progression stages as well as if they are more likely to fall under certain sub-trajectories (see Statistical Annex).

