***Title*** **– Cardiac and vascular disease prevention after hypertensive pregnancy: insights from AI-derived, multi-organ, hypertensive disease progression models**

## *Scientific Importance*

### Pregnancy hypertension, later development of cardiovascular disease and underlying phenotypes

Hypertension during pregnancy [1] occurs in over 10% of women [2] and is associated with up to four-fold increased risk of developing a range of cardiovascular disorders in later life including myocardial infarction, stroke and heart failure [3]. In view of this increase in risk, post-partum review of women who have had a hypertensive pregnancy is advised, followed by annual cardiovascular health assessment [4-9]. However, during this early post-partum period women are often asymptomatic, their blood pressure can temporarily normalise and, currently, there is little personalised advice that can be provided. As a result, loss to follow-up is frequent, particularly in countries without nationalised health care [8] or with limited resources , until the women represents with established hypertension, stroke or myocardial infarction. Indeed, in a UK priority setting partnership of families, researchers and healthcare workers, long-term consequences of pregnancy hypertension, and the best way to prevent them, were identified as two of the most important priorities for future research [10].

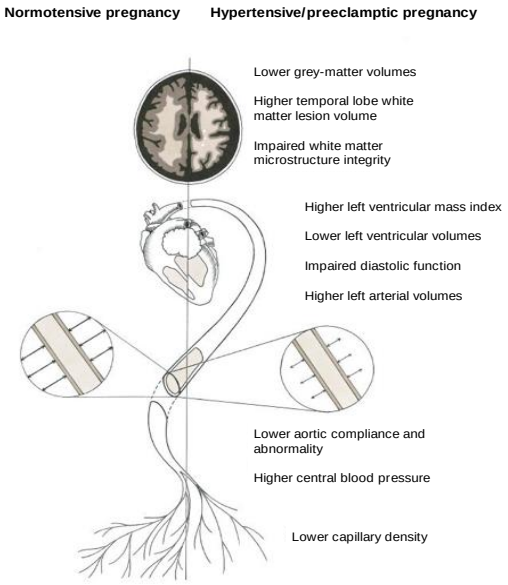


**Figure 1.** Phenotypic changes in the brain, heart, and vasculature in women after a hypertensive pregnancy.

Strong evidence is emerging that during this post-partum period there is identifiable progression of hypertensive-related disease within the heart, brain and vasculature. By undertaking imaging studies in women during the ten years after a hypertensive pregnancy we, and others, have demonstrated left ventricular structure is altered from early after pregnancy and is associated with significant diastolic impairment [11]. These cardiac differences are coupled with later vascular changes extending from central aortic stiffening to peripheral microvascular rarefaction and extend to cerebrovascular small vessel white matter disease [12, 13]. Most of these findings derive from small cohorts of women undergoing single phenotype measures. To address this potential limitation we performed multi-organ, multi-modality imaging to determine whether a panel of imaging markers consistently associated with a hypertensive pregnancy history. This replicated prior single phenotype observations related to cardiac structure, cerebrovascular disease and microvascular rarefaction but, additionally, highlighted that these frequently co-occur and vary independent of blood pressure at time of assessment [13]. When identified in mid-life, the changes we found have, previously, been individually-related to future risk of vascular events independent of blood pressure [14-16].

The longitudinal changes in these patterns of imaging derived phenotypes has not been fully explored. Furthermore, it is unknown whether these phenotypes are also evident in women with different environmental and health-related behaviours in whom advanced disease stages and severe complications are more prevalent and who may gain particular benefit from more accurate disease diagnosis. Children born to hypertensive disorders of pregnancy are, like their mothers, also more likely to develop hypertension in later life as well as being at risk for major events such as stroke [17]. We, and others, have shown they display specific patterns of altered cardiac, cerebral and vascular development similar to those observed in the mothers, in childhood, adolescence, and young adulthood [18, 19]. They are independent of other factors related to the pregnancy of their mother such as degree of prematurity or growth restriction [18] and such findings raise the possibility that unique patterns of hypertensive disease progression may be identifiable across generations of families affected by hypertensive pregnancy.

**Family history of hypertensive pregnancy as a 'model system' to understand and manage early hypertensive disease progression**

Families who have a history of hypertensive pregnancy represent a powerful 'model system' to study early stages of hypertensive disease progression. The new onset of hypertension during a pregnancy identifies both a woman and a child at substantial risk of future problems, who can then be studied over many decades [20, 21]. From these studies, a detailed understanding of the differential rate of hypertension-related change within organs is possible, with the potential to identify factors that drive preclinical hypertensive disease progression and optimal interventions at different disease stages. In addition, optimal approaches to monitor these changes could be developed for clinical application.

During the first decades of life, traditional risk scores, even when modelled as lifetime or relative risk, struggle to provide clinically-meaningful separation of individuals to guide prevention management and provide abstracted information on a potential future event rather than informing on current disease-state [22]. By focusing on underlying disease progression, rather than traditional risk scores, a more personalised life-course approach, targeting prevention from childhood onwards, may be possible to reduce population disease burden [17, 23-26]. However, the breadth and complexity of the information available from biomarkers, often derived at multiple time points, has traditionally presented challenges for both interpretation and simplification into usable clinical tools.

**Integrating multi-modal data into a comprehensive model of hypertensive disease**

Computational modelling and machine learning now allows more sophisticated handling of multi-dimensional data and we, and others, have applied these techniques to various medical and research applications including image processing [27, 28], risk stratification [29], diagnosis [30], prognosis [28], phenotype clustering [31], and treatment selection [32]. Our aim is to find what characterises the hypertensive disease trajectories from large cohorts of cross-sectional data, and our approach is to use contrasted Trajectories Inference (cTI). This approach has shown a great potential in genomics studies [33] and in application to neurodegenerative disease [34], where identified patterns link to neuropathological severity as confirmed by histology as well as memory ability and its future rate of decline [34].

The algorithms in cTI use unsupervised identification and ordering of enriched patterns in a diseased population relative to a comparison background population [34]. Measures that have a high variance across subjects relative to the variance within subjects in the same disease classification group are identified using a contrastive principal component analysis (cPCA). These features are then used to identify patterns, summarised in principal component factors that are present in the diseased population and absent in the healthy subjects. This inherently controls for normal variation in patterns in the healthy population and can deal with the often complicated and non-linear relationships between measures in clinical datasets. From this analysis, sub-trajectories of disease are also identifiable, which, similar to clustering approaches, allows disease subtypes to be identified [35]. In contrast to traditional clustering techniques, this approach can differentiate disease subtypes from different stages of disease.

By applying the cTI approach to echocardiographic features collected in 450 young adults with either normal or elevated blood pressure, we identified preliminary trajectories of change in echocardiographic biomarkers of atrial and ventricular structure and function between 'health' and 'disease' [unpublished data, Figure 2, panel 1a, 2 and 3a]. This is providing novel insight into specific phases of atrial and ventricular coupling and remodelling associated with early hypertension. The work highlights the non-linear nature of both the trajectories and feature associations as well as the power of the technique to capture and display this information. For example, Panel 3aii shows the components of left atrial diastolic function, which is highly weighted in the model, are on divergent trajectories with increasing disease progression while, after an initial increase, left ventricular end-diastolic volume declines. A simple way of characterising this progression is by providing a score for an individual that defines their relative position within the population [Figure 2, panel 4]. We have established that this score, based on resting echo parameters, predicts stress-induced physiological responses in young adults, as well as associates with more complex lifestyle risk scores, previously

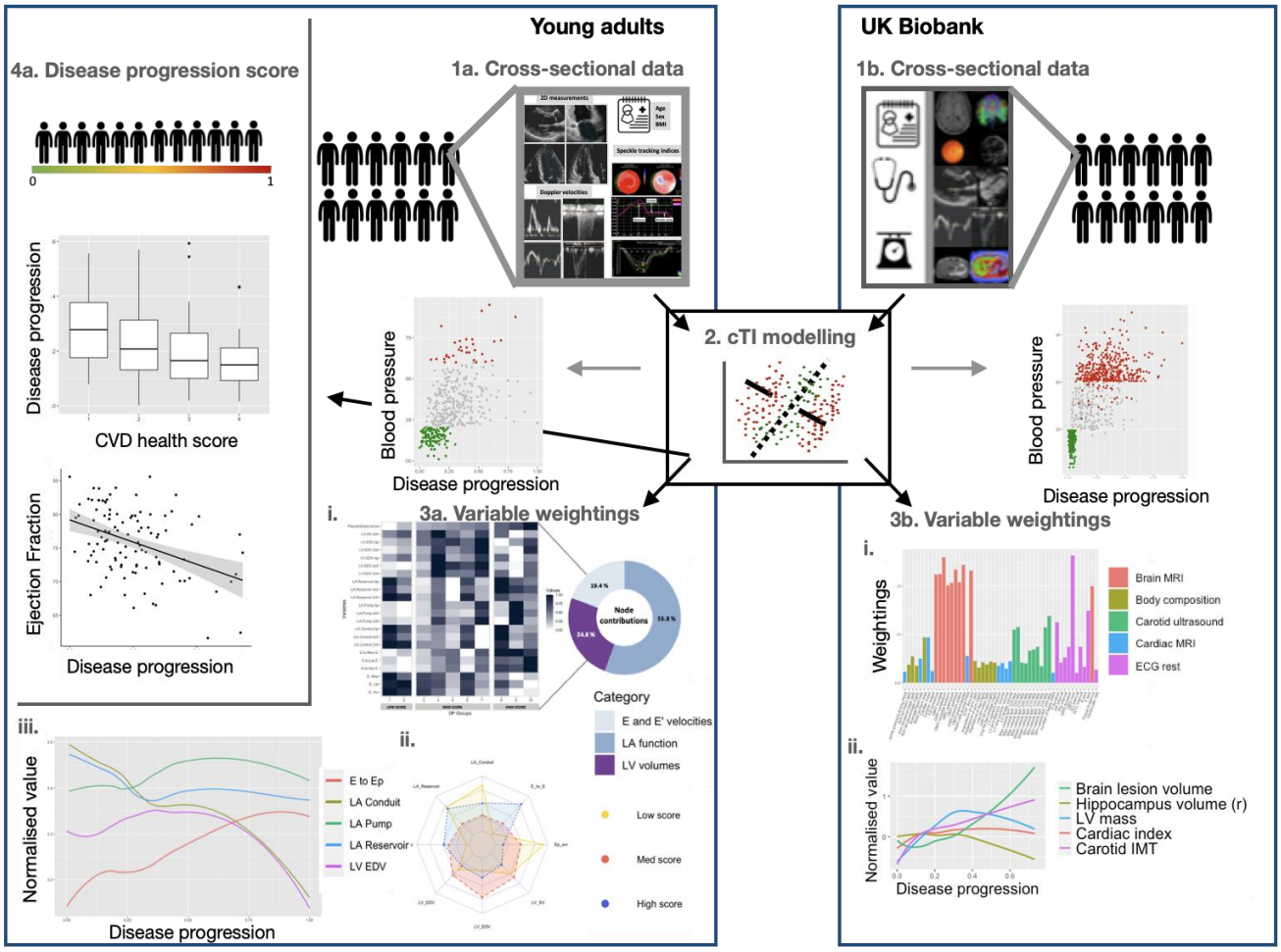


**Figure 2.** Schematic approach for disease progression analysis in (a) young adults and (b) UK Biobank women with hypertension. 1. Multi-dimensional echocardiography data from a cross-sectional cohort (n=411) of young adults and (n=2,200) UK Biobank women with a range of blood pressure measures. 2. The disease progression model was developed using a contrastive Principal Component Analysis (cPCA) algorithm, an unsupervised machine learning tool to identify low-dimensional unique patterns. (2ai/2bi) Separation in disease progression score for healthy as compared to hypertensive participants. 3. Variable weightings show which variables are important in distinguishing the healthy from the hypertensive populations. (3ai) A total of 21 contributed variables were identified and grouped in three categories; left atrial function, left ventricular volumes, and mitral valve E velocity and E' tissue Doppler velocities. (3aii) The mean of highly contributed variables for participants with low, medium, and high scores. (3aiii/3bii). The continuous changes of highly contributed variables throughout the course of the disease progression. (3bi). Variable weightings grouped by modality. (4ai). Participants were assigned into a unique disease progression score from 0 to 1. Lower scores are indicative of pathology free and higher scores indicate advanced stages of hypertension. (4ai) The relationship between disease progression score and a cardiovascular health score based on modifiable risk factors. (4aii) The relationship between disease progression scores and myocardial performance during physical exercise.

shown to identify cerebrovascular disease progression in young adults [16]. In parallel work, we have applied this modelling approach to multi-modality cardiac, brain, ECG and ultrasound imaging data acquired within the UK Biobank Imaging Enhancement, and have been able to summarise biomarker change in transition from health to established hypertensive disease for 2,200 women [Figure 2, panel 1b, 2 and 3b]. Non-linear change across multiple parameters including brain structure, carotid ultrasound and cardiac magnetic resonance are apparent. In addition, using the relative weighting of each variable in determining the disease progression state, a sequential and identifiable pattern for deterioration in individual parameters as hypertensive disease progresses is evident. Panel 3b shows that when including a wide range of biomarkers, brain variables show a strong and consistent pattern in disease progression, with a steady decrease in hippocampal volume and a late rapid increase in brain lesions from moderate to high disease progression. Furthermore, while carotid IMT has a linear increase, the relationship with left ventricular mass is more complex with a decline between moderate and high disease progression scores after an initial phase of remodelling.

**Creating clinical tools from multi-organ imaging**

An advantage of this approach is that the combination of imaging features present for an individual can be used to determine their relative position on the identified path of disease progression, or their specific trajectory of disease where different subtype trajectories are identified [36]. In hypertensive disease, such information provides a more meaningful, instantaneous assessment of disease burden for an individual, over and above what can be interpreted from a blood pressure measure, which may be transiently influenced by environment or medication. To aid interpretation and communication of this disease progression for a patient, the relative position for that individual can be presented to them as a score. Also, by understanding how imaging measures contribute to different score levels, insights can be generated into whether there are stages of disease progression in different parts of the body that may benefit from specific management strategies.

A potential limitation of this approach is that clinical translation may remain limited to a few facilities if information needs to be derived from complex measures since, in clinical settings, including within the UK, such phenotypic data is not available for most patients [37]. Besides, effective machine learning systems also need to take account of potential variation in patterns associated with a disease across geographical or genetic differences if they are to be generalizable worldwide. In order to tackle these two limitations, our approach is to acquire 'affordable' measures and map them into our complex and rich model. 'Affordable' measures derived from ECG, echocardiography and retinal imaging are already routinely performed in current clinical practice around the world to detect end organ changes in cardiovascular disease and only a small proportion of the rich information they contain is used [37, 38]. By collecting both affordable and deep phenotyping parameters for an individual, the affordable parameters can be modelled against the features in the more complex deep phenotype model. A disease progression can be calculated form the resulting affordable model, with the potential for clinical use [34].

***Objectives for research plans***

Our primary aim is to understand hypertensive disease progression over the life course of women and their children after a hypertensive pregnancy. We will use this information to address our secondary aim, which is to develop clinical approaches to prevent early onset cardiac, cerebral and vascular disease in these families. These aims will be achieved through delivery of these specific objectives:

1) Studies in women
    a) Collate longitudinal complex imaging data across brain, heart and vasculature in women from 6 months to 25 years post-pregnancy from existing and new data collections.
    b) Describe longitudinal changes in organ phenotypes from pregnancy to later life related to hypertensive pregnancy.
2) Studies in young adults
    a) Collate complex imaging data across brain, heart and vasculature in young adults born to hypertensive pregnancies using existing and new data collections.
    b) Describe differences in organ phenotypes between young adults born to a hypertensive and normotensive pregnancy.
3) Computational disease progression models
    a) Build computational models that combine multiple imaging data sources to describe phenotypes related to a hypertensive pregnancy in women and young adults.
    b) Build computational models to describe hypertensive disease progression in UK Biobank imaging data including sub-trajectories of hypertensive disease.
    c) Compare disease progression related to hypertensive pregnancy in women and young adults to patterns of hypertensive disease progression defined in UK Biobank.
4) Clinical translation
    a) Develop methods, such as scores, for summarising disease progression described by complex imaging measures defined in computational models.
    b) Identify an optimal combination of clinically relevant measures that recapitulate the complex scoring system
    c) Investigate whether the clinical scoring systems can predict adverse cardiovascular outcomes and are modifiable.

***Potential added value***

This programme of work is focused on applications in hypertensive pregnancy. It will establish patterns of disease development related to hypertension and will develop modelling work to unravel these progressive changes in multiple organs. This has direct applications in other areas of hypertension research and provides new insights in the progression of subclinical to critical disease stages. The expertise we are developing in cardiovascular disease progression could have direct relevance to application in other clinical datasets and generate insights from currently unused data. The availability of additional data sources such as blood and urine samples in these cohorts will also create a platform for collaborative work at multiple system levels including linking future genomic and metabolomic knowledge with our organ-specific disease progression models. The relative importance of the measures in hypertensive disease progression will allow development toward more accurate disease monitoring tools to generate efficiencies within healthcare practice. In the future, this could be extended to include measures that can be acquired outside the hospital setting even used directly by the patient to empower them to monitor and manage their own health.

***Scientific potential***

**Environment -** This programme of work will be based in Oxford but brings together researchers from multiple centres. It builds on over a decade of collaborative projects amongst the lead groups that has combined clinical trials with imaging and computational modelling to tackle long term risks associated with hypertensive pregnancy. These groups have been behind key publications describing the development of cardiovascular risk in mothers and offspring of hypertensive pregnancy from childhood through to later life [6, 11-13, 18, 25, 39-46]. In addition, we have pioneered computational modelling of the heart, brain and vasculature to unravel key phenotypic variations that emerge after hypertensive pregnancy [12, 13, 18]. Furthermore, by adopting rapid developments in machine learning approaches and working collaboratively with key engineering groups we have been able to develop innovative image processing technologies that have successfully translated into clinical practice [47]. The experimental plans rely on existing archived datasets and new clinical studies. The existing datasets have been compiled over a 15-year period and are accessible because of a cloud-based image and database trusted research environment that has been developed as a bespoke construct for the research group in Oxford. This allows rapid deployment of relevant computational approaches with easily accessible datasets. In addition, the investigators developed key imaging protocols applied within the UK Biobank Imaging Enhancement and advise on the delivery of this UK Biobank project. This privileged insight into the UK Biobank Imaging Enhancement will ensure the optimal use and application of available imaging datasets. Furthermore, this link means imaging protocols applied within UK Biobank have been used within the studies performed in Oxford for many years ensuring a consistency of imaging data across groups. Three new datasets are planned to enhance the breadth and complexity of the available imaging data drawn from Oxford, Bristol (ALSPAC) and South Africa. Collection of these datasets is possible because of access to established world-leading imaging and clinical trial units in the UK and South Africa led by the principal and co-investigators in this programme grant. This programme therefore brings together a strong team focused on clinical translation of multi-organ imaging and computational modelling to prevent long-term disease associated with hypertensive pregnancy.

**Investigator group -** The investigator group brings together researchers in Oxford, London, Bristol, Montreal, and Cape Town who have worked together for over 10 years. The research leads include: Paul Leeson, who has a strong focus and corresponding publication record on using imaging and laboratory studies to identify early cardiac and vascular changes in young people at risk of cardiovascular disease and pioneering use of artificial intelligence to improve the clinical tools available to identify those at risk; Adam Lewandowski, who is a BHF Intermediate Research Fellow, working at the forefront of research in the field of cardiovascular remodelling and systems physiology in people born to pregnancy complications, unravelling changes in cellular and physiological pathways; Winok Lapidaire, who is the Lee Hysan Research Fellow at St Hilda's College, Oxford with a multidisciplinary background in neuroimaging, medical sciences, and data science. This core group is supported by expertise across three major domains. ***(1) Hypertensive pregnancy,*** Lucy Chappell is a leader in the field of Obstetrics and long term cardiovascular risk management who is influential in the pregnancy hypertension research and development of clinical guidelines in this area; Basky Thilaganathan has led conceptual thought in the area of cardiac remodelling after

hypertensive pregnancy and published seminal papers describing long term phenotypic heart failure disease patterns; Abigail Fraser is an ALSPAC co-investigator and has worked extensively to understand the relationship between pregnancy complications and women's cardiovascular health with a programme of work using electronic health records to track life course trajectories of cardiovascular risk in women with pregnancy complications; Ntobeko Ntusi leads a cardiovascular imaging research programme in Cape Town combining complex phenotyping measures alongside simple measures in the low resource setting, working closely with collaborators in Obstetrics who successfully deliver clinical trials and studies in diverse settings. ***(2) Clinical translation,*** Christina Aye has close insight of the practicalities of management of hypertensive pregnancy and delivery of clinical studies; Lucy MacKillop has pioneered development of clinical tools for management of risk in pregnancy and beyond that take simple measures for application into clinical practice; Richard McManus has been world leading in the field of self- monitoring in hypertension and leads active programmes of work related to management of hypertension in pregnancy. ***(3) Artificial intelligence and modelling,*** Eric Ohuma has expertise in longitudinal cardiovascular modelling through fetal life having worked on some of the largest pregnancy imaging studies globally; Pablo Lamata is a conceptual thought leader in the sphere of computational modelling particularly within cardiology and a pioneer of development of digital twin models; Yasser Itturia Medina leads the field of application of disease progression models to neurodegenerative disease and has guided the translation of the computational methodological approaches from neuroscience to cardiovascular disease; Ana Namburete is a leader in the field of neuroimage analysis, particularly ultrasound, and specialises in data harmonisation in large imaging studies to facilitate computational algorithm development.

## ***Research plans***

**General overview -** The research programme will be delivered by a dedicated project manager and associated research staff under the direction of the Principal Investigators: Paul Leeson, Adam Lewandowski and Winok Lapidaire. The project manager will be responsible for monitoring progress of the workstreams against defined targets to ensure dependencies are managed appropriately to maintain progress against the planned 5-year timeline [GANNT chart available on request]. The investigator group have refined the research objectives and will guide ongoing direction of the workstreams through regular investigator meetings. Individual investigators will provide intellectual and practical contributions to ensure timely progress of each workstream within their area of expertise. Four workstreams are planned: the first will collate, collect and harmonise imaging data from women six months to 25 years after pregnancy and use standard statistical approaches to describe longitudinal variation within paired samples and cross sectional differences between pregnancy groups; the second will collate, collect and harmonise imaging data in both male and female young adults and undertake standard comparisons of differences in phenotypes according to the maternal pregnancy history; and the third will utilise the harmonised datasets to build computational models to describe phenotype variation and progression related to hypertensive pregnancy in young adults and women post-pregnancy. In addition, this workstream will compare disease progression to a more comprehensive hypertensive disease progression model developed from UK Biobank Imaging Enhancement datasets. Finally, the fourth workstream will

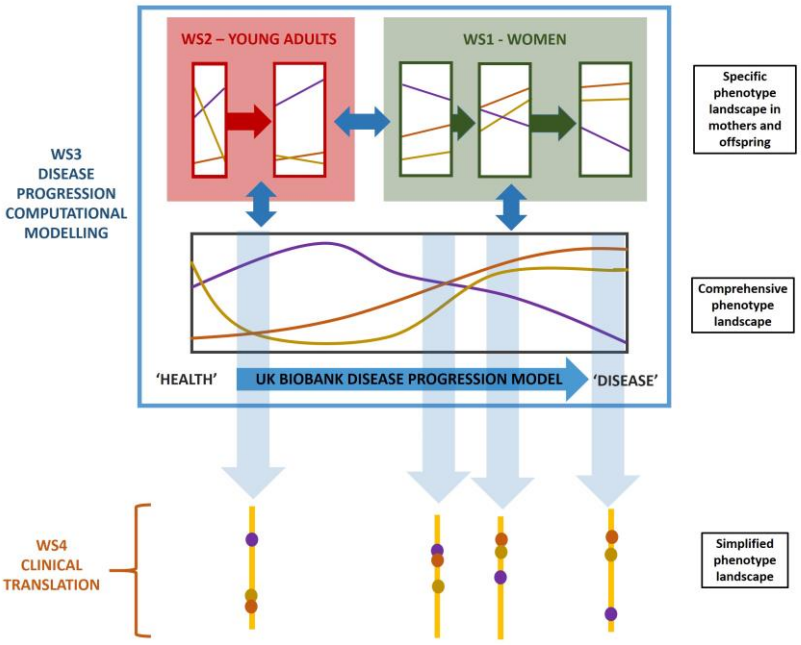


**Figure 3.** Schematic demonstration of the four planned workstreams.

identify clinically translatable approaches to define disease stages from the model datasets and scoring systems based on both complex imaging parameters and accessible clinic-based measures.

**Datasets and imaging protocols –** All existing Oxford datasets have followed the same data acquisition following standard operating procedures and data processing protocols, which are in line with the UK Biobank pipeline [48]. Imaging protocols include MRI (cardiac, brain, liver, aorta), echocardiography, retinal imaging, blood sample collection for both women and young adults. MR data is acquired on a 3T Siemens Prisma MR scanner for POP-HT, CAREFOL-HT, PVS-follow-up, and #SELFIE (see Table 1 and 2 for study descriptions). YACHT and TEPHRA cohorts were scanned on a 3T Siemens Trio, and PVS and EVS participants on a 1.5T Siemens Sonata. All new datasets will be collected using the same protocols on 3T Prisma scanners. An electronic template for databases is housed in Castor EDC, an online database management system, to ensure consistent variable labelling. Calculations are integrated in the study template and performed from source data to reduce human error (see Data Management Plan for details). The ComBat analysis package will be used for combined datasets to limit location and scale effects due to acquisition and time differences [49, 50]. This software is consistently shown to remove unwanted sources of variability, specifically site differences, while preserving meaningful biological variation. ComBat uses the premise that location and scale effects for multivariate outcomes are drawn from a common parametric prior distribution, then estimates hyperparameters of prior distributions using an empirical Bayes framework to compute conditional posterior estimates of all effects.

**Workstream 1 (Lead: Leeson): Develop datasets and describe disease progression in women after hypertensive pregnancy to 25 years post-partum**

***Workstream 1.1: Integrate and harmonise datasets from existing post-pregnancy cohorts (Table 1) -*** Existing (or currently being acquired) imaging datasets involving 600 women after either a normotensive or hypertensive pregnancy are available and extend from the first year (POP-HT, CAREFOL-HT) to 5 to 13 years post-partum (PVS). Within this workstream we will ensure standardisation of legacy data and imaging variables from our previous studies with our recent and new planned data collections. Missing variables will be identified and, if necessary, additional metrics extracted from archived imaging datasets to maximise data completeness and limit missingness within the combined dataset.

***Workstream 1.2: Collect new data in women 20 to 25 years after pregnancy -*** We will study a group of 200 women 20 to 25 years post pregnancy to extend data coverage over time and age to ensure broad representation of hypertensive disease state during mid-life in women. This will be achieved by follow-up of participants from our previous Preeclampsia Vascular Study (PVS) originally seen at 6 to 13 years post-index hypertensive or normotensive pregnancy. If required, we will restart our identification and recruitment methods, previously used in PVS, based on obstetric records to identify and recruit additional women who had pregnancies 10 to 20 years prior, to ensure sample size is reached. All participants will undergo a standardised imaging protocol as described in the general overview.

***Workstream 1.3: Collect new data in women from different demographic backgrounds -*** To increase heterogeneity within the dataset we will recruit a new cohort of 200 women at six months post-partum in partnership with our co-investigators in Cape Town, South Africa. This will allow us to investigate patterns of disease in women who have had a hypertensive pregnancy in a primarily non-white population and low-income setting. Furthermore, it will provide insight into the potential variation in disease patterns globally. Recruitment of the women for the study will be done by the Department of Obstetrics and Gynaecology team at the Groote Schuur Hospital, which is a public healthcare hospital in Cape Town. Imaging will be performed at the neighbouring Cape Universities Body Imaging Centre (CUBIC). Standardised imaging protocols will be used as performed in Oxford to allow integration of the datasets.

***Workstream 1.4: Study changes in phenotypes of women up to mid-life after pregnancy -*** Using the harmonised datasets from 6 months to 25 years post-partum we will undertake standard descriptive and statistical analysis of both cross-sectional differences and longitudinal variation in phenotype markers, relative to time from index pregnancy. As this analysis will include women who had normotensive pregnancy, this will allow us to establish mean levels and variation in measures in women at different ages for a range of cardiac, brain and vascular indices post-pregnancy. This

allows development of a nomogram against which deviation in specific hypertensive pregnancy groups can be compared. Differences will be reported at single time points according to demographic background and pregnancy history, taking into consideration contemporary definitions of hypertensive pregnancy and subtypes including late and early onset preeclampsia and gestational hypertension (see Statistical Annex). Furthermore, we will be analyse paired data for individuals in PVS to study rate of progression over time taking into account lifestyle and cardiovascular risk assessed at baseline and follow-up using regression and interaction analysis.

**Table 1: Datasets in women (new datasets to be collected within programme shaded grey)**

| Study | Description | Sample size |
|---|---|---|
| Six months post partum | | |
| POP-HT | Randomised control trials of blood pressure management in hypertensive pregnancy post-partum period | 200 hypertensive pregnancy,<br>50 normotensive pregnancy |
| CAREFOL-HT | Randomised control trial of prenatal nutritional folic acid supplement in hypertensive pregnancy | 100 hypertensive pregnancy,<br>50 normotensive pregnancy. |
| South Africa | Observational study of women with and without hypertensive pregnancy with high and low income in Cape Town | 100 hypertensive pregnancy group,<br>100 control group |
| 5-10 years post-partum | | |
| PVS | Observational study of women with and without hypertensive pregnancy | 164 hypertensive pregnancy group,<br>75 normotensive pregnancy group |
| 20-25 years post-partum | | |
| PVS follow-up | Observational study of women with and without hypertensive pregnancy | 100 hypertensive pregnancy group,<br>100 normotensive pregnancy group |

***Workstream 2 (Lead: Lewandowski): Develop datasets and define disease in offspring of hypertensive pregnancy compared to normotensive pregnancy***

***Workstream 2.1: Harmonise datasets from existing young adult cohorts (Table 2) -*** Existing datasets from our studies include 710 young adults born either to a normotensive or hypertensive pregnancy. The cohorts are enriched for preterm birth to allow study of relative contributions of prematurity and hypertensive pregnancy to disease development. We will ensure consistency of variable names across studies based on our current standardised methods. Where necessary additional metrics will be extracted from archived imaging datasets to ensure data completeness and consistency across datasets.

***Workstream 2.2: Collect new data in young adults born to hypertensive and normotensive pregnancies -*** To expand our data availability in this age group we will undertake a follow-up of 200 young adults from the Avon Longitudinal Study of Parents and Children (ALSPAC) study, of which 100 will have been born to a hypertensive pregnancy and 100 following an uncomplicated pregnancy. At time of follow-up, they will be 30 to 35 years of age. At the 25-year follow-up in Bristol, there were 610 individuals born to hypertensive pregnancies. Participants will be invited by the ALSPAC study team to attend a study visit in Oxford to perform MRI of the brain, heart and vasculature using our standardised protocols. The data collection will also include all measures used in the Oxford studies including assessment of cardiac function in response to physiological stress, both using echocardiography and MRI. A dataset including information from previous follow-up of these cohorts [20] and the detailed pregnancy data available within ALSPAC will be generated and integrated with available data from the Oxford cohorts.

***Workstream 2.3. Compare imaging phenotypes in young adults born to either hypertensive or normotensive pregnancy -*** Within the Oxford and ALSPAC datasets we will undertake simple comparisons of different imaging phenotypes according to pregnancy history to understand the

extent of disease progression during young adult life. Although the specific comparison will be between those born to a hypertensive pregnancy and those born to a normotensive pregnancy (see Statistical Annex), we will additionally study whether these associations vary with features of pregnancy history such as gestational age at delivery and degree of growth restriction. Through standard regression analysis we will study the association of pregnancy and current lifestyle factors on disease progression within the heart, brain and vasculature during young adult life. Furthermore, we will attempt to determine how phenotype varies over the age range of young adulthood between age 18 and 40 years of age. As multiple datasets will be available, we will be able to test whether differences identified in one study can be replicated in independent datasets.

**Table 2: Young adult datasets (new datasets to be collected within programme shaded grey)**

| Study | Description | Sample size |
|---|---|---|
| Young adults aged 18 to 40 years | | |
| EVS/TEPHRA/ YACHT/#SELFIE | Young adults with normal and elevated blood pressure, including after pregnancy complications resulting in preterm delivery and/or hypertensive pregnancy | 50 hypertensive pregnancy offspring group, 660 normotensive pregnancy offspring group |
| Young adults aged 30-35 years | | |
| ALSPAC | Young adults recruited from ALSPAC cohort, including both normotensive and hypertensive pregnancy offspring | 100 hypertensive pregnancy offspring group, 100 normotensive pregnancy offspring group |

**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

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).

**Table 3: Study investigation per combined dataset**

| Datasets | | Cardiac/Aortic MRI | | | | Brain MRI | | | | | Abdominal MRI | | | Echocardiography | | | | | | CU | RI | DEM | Physical Measures | | | Blood Pressure | | ECG | PWA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | - Short-axis cine stack | - Long-axis cine stack (horizontal, vertical, left ventricular outflow) | - T1 mapping | - Aortic distensibility | - T1 | - T2 | - DWI | - ASL | - TOF | - T1 | - T2* | - DIXON | - 2D | - Colour Doppler Flow | - Continuous Flow Doppler | - Motion Mode | - Pulse Wave Doppler | - Tissue Doppler Imaging | | | | - Anthropometry | - Body Composition | - Spirometry | - Seated | - Ambulatory | | |
| YA | F M | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| W | UK SA | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| BB | F M | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | | | | | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*CU=Carotid Ultrasound; RI=retinal imaging; DEM-demographic; ECG=electrocardiogram; PWA=pulse wave analysis; YA=Young adults; W=women; BB=UK Biobank; UK=United Kingdom; SA=South-Africa; F=female; M=male*

## ***Workstream 4 (Leads: Leeson, Lewandowski, Lapidaire): Identify optimal clinical translation***

***Workstream 4.1: Develop optimal methods for calculation of a HyperScore based on complex measures to summarise disease progression -*** As cTI calculates relative position of an individual within the principal component space, using the Euclidean Distance Matrix, the proximity of each participant to the healthy population can be calculated with a Minimum Spanning Tree. Therefore, a score can be generated for any participant that, in our pilot work, we have termed the “HyperScore”. In this workstream we will study the practicalities of generation of a HyperScore for an individual patient. Testing and prototype work will make use of individual datasets drawn from our existing data resource of women and young adults (see Statistical Annex). We will study optimal methods for score generation including, as example, identification of the minimum datasets required for stable score generation and the practicalities of calculation of a continuous score, threshold or range. In addition, we will identify optimal interfaces/programmes for new data entry to generate an output for an individual.

***Workstream 4.2: Identify whether single or a combination of simple phenotypes measures can recapitulate the complex HyperScore -*** The UK Biobank and our own datasets contain complex data that would not normally be available within a healthcare system. Therefore, in this workstream we will work at modelling a HyperScore that recapitulates the complex HyperScore developed in workstream 4.1 but uses only affordable measures that are currently available in clinic. To undertake this work, we will use our combined datasets from Workstream 1 and 2 that contain both deep phenotyping and affordable measures and include echocardiography, ECG and retinal images. To develop the affordable score, we will identify the optimal combination and weighting of affordable measures that approximate to the reference HyperScore as the ground truth. We will test its accuracy as compared to the deep phenotyping HyperScore in the UK and South Africa cohorts. To develop the minimum viable score, we will also test how effectively the accuracy of the affordable Hyperscores is maintained as one or more modalities are excluded (see Statistical Annex).

***Workstream 4.3: Assess whether single measures, complex HyperScore or simple HyperScore can be predict longer-term outcomes and could be modified by interventions. -*** In the last workstream, we will undertake three distinct pieces of work. Firstly, we will study association between longitudinal clinical progression in groups of individuals and the individual imaging features, the complex HyperScore and simplified HyperScore. Specifically, if available at sufficient scale by the end of the programme grant, we will make use of clinical event outcome data in UK Biobank to understand association between scores and clinical events. In addition, we will use longitudinal data from the PVS follow-up to study associations with progression over shorter time periods. The second piece of work will be to understand modifiability of disease progression by looking at impact of interventions that have been undertaken within our peri-partum randomised trials (CAREFOL-HT, POP-HT) on disease progression (see Statistical Annex). In the future, we will use our collaborations with the broader research community to understand whether we can replicate and validate our scoring systems in other datasets including (as detailed in attached letters of support) the Queen of Heart's study, Generation R, and iPlacenta.

**Exploitation and dissemination -** All results will be disseminated through investigator meetings and academic conferences to ensure wide awareness within the research community. In addition, a series of high impact academic publications are expected to be forthcoming. Patient groups have been involved in the development of this proposal and will ensure dissemination of our findings widely to the public. The principal investigators and co-investigators are heavily involved in research and guideline development within national and international bodies for cardiovascular prevention and obstetrics. Therefore, it should be possible to define a clear route to clinical translation and acceptance if clinical validation is promising. As regard exploitation, we are in ongoing discussion with Oxford University Innovation and the Translational Research Office about the best ways to translate the computational modelling of disease progression into a usable tool. To do this we have filed intellectual property on the 'HyperScore' as an approach to stage disease development within a hypertension population. During the programme of work we will explore routes to commercialisation/translation including whether a HyperScore model can be created based on measures that are increasingly being performed through self-monitoring in a home or community environment. We envisage multiple translational routes for HyperScore models as health applications; in clinic and/or by patients themselves, either integrated or independently.

**Ethics and research governance -** Local clinical research and ethical governance teams will oversee all projects. Although personal information will be accessed within clinical studies, staff will ensure participant anonymity is maintained through use of study specific participant ID for all assessments and documents e.g. CRF and electronic database, which will be password protected. Any personal identifying information (such as contact details) will only be stored for the duration of ethical approval and only on local site High Compliance Servers. Only authorised team members will be able to access these servers. Once practical, links between anonymised and identifying information will be broken, consistent with GDPR/Data Protection Act 2018. Any paper documents will be stored in locked cupboards in restricted access areas, only accessible by study staff and authorized personnel. Studies will be registered and regularly updated on the ISRCTN registry. Findings from the workstreams and sub-sections will be widely disseminated to academic audiences to maximise potential academic impact. Results will be presented at national and international scientific meetings and published in full in peer-reviewed journals as Open Access papers only. Due to the nature of the workstreams and timelines for the deliverables, we expect publications to begin within the first year of the programme and continue throughout the five-year period. Where appropriate, following discussion between the investigators, access to data collected and analysed as part of the research programme may be provided to other interested parties.

## *REPRODUCIBILITY AND STATISTICAL DESIGN ANNEX*

**Experimental approach and power calculations for workstreams 1 and 2**

*Workstream 1.4:* Analysis of covariance (ANCOVA) will be used to assess cross-sectional group differences and power has been estimated based on left ventricular mass/end diastolic volume (LVM/EDV), adjusted for age; known to differ between women after hypertensive pregnancy (HP) and normotensive pregnancy (NP) in the year after pregnancy. Differences will be studied at 6 months, 5-10 years, and 20- 25 years post-pregnancy. Sample size calculations are at 90% power for a two-sided test at α=0.05 using the following formula: $n_i = 2(\frac{Z_{1-\frac{\alpha}{2}}+Z_{1-\beta}}{ES})^2$.

To detect a 0.32g/ml LVM/EDV difference, in line with a conservative estimate of differences in young adults with different pregnancy background [51], we require 19 participants per group. The planned sample sizes of 380 HP vs 150 NP (6 months), 164 HP vs 75 NP (5-10 years), 100 HP vs 50 NP (20-25 years) are thus sufficient.

*Workstream 2.3:* ANCOVA will also be used in the young adults and power has been estimated for differences in right ventricular mass index (RVMI), adjusted for age and sex, as previously shown to differ in infants born to HP. There is no data on RVMI in adults born to HPs, but infants born to HPs have a 23% difference in RVMI at 3 months postnatal age [52] and a similar RVMI difference remains from infancy to adulthood in people born preterm, who have high rates of HP history [18, 52]. Detection of a highly conservative 10% mean difference, based on term born young adult reference values (20.4±3.4 g/m$^2$) [18], requires n=59 in each group. The planned sample size of 150 HP vs 760 NP young adults is more than sufficient.

An interaction term will be added to secondary analysis on subtypes including late and early onset preeclampsia and gestational hypertension. Secondary outcomes will be left ventricular (LV) mass indexed to body surface area, LV wall thickness, LV diastolic function, white matter hyperintensities, and retinal vessel thickness.

**Experimental approach for workstreams 3 and 4**

For each dataset, variables with >50% missing data and subjects with >20% missing data are excluded. Remaining missing data will be imputed using trimmed score regression (TSR)[53].

*Workstream 3.3:* A proportion test will be used to identify significant differences in the proportion of HP women, South-African HP women, and young adults born to HP across sub-trajectories as compared to the proportions of subjects in the UK Biobank across these sub-trajectories.

*Workstream 4.1:* A five-fold cross validation approach will be used for stability testing and internal validation with datasets split 80% training and 20% testing and a representative proportion of hypertensive and normotensive participants. Adequate stability requires limited effects of removing small proportions of datapoints from training data, i.e. the mean variation of the training HyperScores in relation to the full data HyperScore to remain below one standard deviation. Internal validation requires no statistical difference between the proportions of subjects in the wrong HyperScore bracket (i.e. a normotensive above the hypertensive threshold or a hypertensive below the normotensive threshold) in training and testing subjects. The normotensive threshold is the value below which <5% of hypertensive subjects are placed and the hypertensive threshold the value above which <5% of normotensive subjects are placed in the training dataset.

*Workstream 4.2:* The UK women dataset will be split 80% training and 20% testing. The affordable HyperScore model will be trained on training subjects and applied to the testing data subjects. In these testing subjects, the affordable HyperScores will be compared against the deep phenotype HypersScore using a calibration (c-index) and discrimination index. We will repeat this process in the South Africa data. To develop the minimum viable score, we will repeat the previous steps multiple times with all different combinations of the following modalities; cardiac MRI, brain MRI, liver MRI, carotid ultrasound, echocardiography, anthropometry and blood pressure measures. Affordable HyperScores that are not significantly different from the deep phenotype HyperScore will be deemed viable.

*Workstream 4.3:* Using an ordered logistic regression, deep phenotype and affordable HyperScores will be assessed for association to the risk of stroke or heart attack in the UK Biobank (only deep phenotype HyperScore) and with high blood pressure, stroke or heart attack in PVS. Pregnancy interventions will be assessed using a mixed effects model with baseline value, minimisation factors used in the randomisation process, randomised group and time fitted as fixed effects with a random intercept for each participant and HyperScore as the outcome variable.

***References***


1. Robertson, R.M., Circulation, 2001. **103**(19): p. 2318-20.
2. Ros, H.S., S. Cnattingius, and L. Lipworth, Am J Epidemiol, 1998. **147**(11): p. 1062-70.
3. Bellamy, L., et al., BMJ, 2007. **335**(7627): p. 974.
4. Mosca, L., et al., Circulation, 2007. **115**(11): p. 1481-501.
5. Mosca, L., et al., Circulation, 2011. **123**(11): p. 1243-62.
6. Lazdam, M., et al., Hypertension, 2012. **60**(5): p. 1338-45.
7. Moe, K., et al., Eur J Prev Cardiol, 2019: p. 2047487319879791.
8. Lui, N.A., G. Jeyaram, and A. Henry, Front Cardiovasc Med, 2019. **6**: p. 160.
9. Gamble, D.T., et al., Front Cardiovasc Med, 2019. **6**: p. 55.
10. Ho, A., et al., BMJ Open, 2020. **10**(7): p. e036347.
11. Melchiorre, K., et al., Hypertension, 2011. **58**(4): p. 709-15.
12. Siepmann, T., et al., Neurology, 2017. **88**(13): p. 1256-1264.
13. Boardman, H., et al., Hypertension, 2019. **(in press)**.
14. Ghossein-Doha, C., et al., Hypertension, 2013. **62**(2): p. 382-90.
15. Piepoli, M.F., et al., Eur Heart J, 2016. **37**(29): p. 2315-81.
16. Williamson, W., et al., JAMA, 2018. **320**(7): p. 665-673.
17. Fox, R., et al., Journal of clinical medicine, 2019. **8**(10): p. 1625.
18. Lewandowski, A.J., et al., Circulation, 2013. **127**(2): p. 197-206.
19. Davis, E.F., et al., BMJ Open, 2015. **5**(6): p. e008136.
20. Boyd, A., et al., Int J Epidemiol, 2013. **42**(1): p. 111-27.
21. Fraser, A., et al., Int J Epidemiol, 2013. **42**(1): p. 97-110.
22. Hanssen, H., et al., Eur J Prevent Cardiol, 2021. **(in press)**.
23. Lewandowski, A.J., et al., Hypertension, 2020. **76**(4): p. 1028-1037.
24. Cairns, A.E., et al., Hypertension, 2018. **72**(2): p. 425-432.
25. Bramham, K., et al., Bmj, 2013. **346**: p. f894.
26. Horvath, S. and K. Raj, Nat Rev Genet, 2018. **19**(6): p. 371-384.
27. Oikonomou, E.K., M. Siddique, and C. Antoniades, Cardiovascular Research, 2020.
28. Leiner, T., et al., Journal of Cardiovascular Magnetic Resonance, 2019. **21**(1): p. 61.
29. Goldstein, B.A., A.M. Navar, and R.E. Carter, European Heart Journal, 2017. **38**(23): p. 1805-1814.
30. Sanchez-Martinez, S., et al., Circ Cardiovasc Imaging, 2018. **11**(4): p. e007138.
31. Katz, D.H., et al., Journal of Cardiovascular Translational Research, 2017. **10**(3): p. 275-284.
32. Cikes, M., et al., Eur J Heart Fail, 2019. **21**(1): p. 74-85.
33. Magwene, P.M., P. Lizardi, and J. Kim, Bioinformatics, 2003. **19**(7): p. 842-50.
34. Iturria-Medina, Y., et al., Brain, 2020. **143**(2): p. 661-673.
35. Kane, G.C., et al., JAMA, 2011. **306**(8): p. 856-863.
36. Kane, G.C., et al., Jama, 2011. **306**(8): p. 856-63.
37. https://www.nice.org.uk/guidance/ng136/chapter/Recommendations#monitoring-treatment-and-blood-pressure-targets, Accessed 23/10/2020.
38. Augustine, D., et al., Echocardiography, 2015. **32**(2): p. 302-9.
39. Cairns, A.E., et al., BMJ open, 2017. **7**(11): p. e018696-e018696.
40. Cairns, A.E., et al., Hypertension, 2018. **72**(2): p. 425-432.
41. Lazdam, M., et al., J Pregnancy, 2012. **2012**: p. 704146.
42. Lazdam, M., et al., Hypertension, 2010. **56**(1): p. 159-165.
43. Melchiorre, K., et al., Hypertension, 2011. **57**(1): p. 85-93.
44. Timpka, S., et al., Journal of the American Heart Association, 2016. **5**(11): p. e003906.
45. Yu, G.Z., et al., Hypertension (Dallas, Tex. : 1979), 2016. **68**(3): p. 749-759.
46. Yu, G.Z., et al., Hypertension (Dallas, Tex. : 1979), 2018. **72**(4): p. 937-945.
47. Mumith, A., et al., EuroEcho Imaging conference, Vienna, Austria (in press). 2019.
48. Littlejohns, T.J., et al., Nat Commun, 2020. **11**(1): p. 2624.
49. Johnson, W.E., C. Li, and A. Rabinovic, Biostatistics, 2007. **8**(1): p. 118-127.
50. Pomponio, R., et al., Neuroimage, 2020. **208**: p. 116450.
51. Huckstep, O.J., et al., Journal of the American College of Cardiology, 2018. **71**(12): p. 1347-1356.
52. Aye Christina, Y.L., et al., Journal of the American Heart Association, 2020. **9**(9): p. e014586.
53. Folch-Fortuny, A., F. Arteaga, and A. Ferrer, Chemometrics and Intelligent Laboratory Systems, 2016. **154**: p. 93-100.