Cardiac magnetic resonance radiomics reveal differential impact of sex, age, and vascular risk factors on cardiac structure and myocardial tissue

Zahra Raisi-Estabragh[1,2†], Akshay Jaggi[4†], Polyxeni Gkontra[4], Celeste McCracken[1,3], Nay Aung[1,2], Patricia B. Munroe[1], Stefan Neubauer[3], Nicholas C. Harvey[5, 6], Karim Lekadir[4], Steffen E. Petersen[1,2, 7, 8*]

1. William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK
2. Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, West Smithfield, EC1A 7BE, UK
3. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, OX3 9DU, UK
4. Departament de Matemàtiques & Informàtica, Universitat de Barcelona, Spain
5. MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton, SO16 6YD, UK
6. NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, SO16 6YD, UK
7. Health Data Research UK, London, UK
8. Alan Turing Institute, London, UK

[†]These authors have contributed equally and share first authorship

*Corresponding author: Professor Steffen E. Petersen. William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK; Email: s.e.petersen@qmul.ac.uk; Telephone: +44-2078826902

Running title (5 words): vascular risk factors radiomics phenotypes

Data availability statement

This research was conducted using the UKB resource under access application 2964. UK Biobank will make the data available to all bona fide researchers for all types of health-related research that is in the public interest, without preferential or exclusive access for any persons. All researchers will be subject to the same application process and approval criteria as specified by UK Biobank. For more details on the access procedure, see the UK Biobank website: http://www.ukbiobank.ac.uk/register-apply/.

Funding statement

This project was enabled through access to the MRC eMedLab Medical Bioinformatics infrastructure, supported by the Medical Research Council (www.mrc.ac.uk; MR/L016311/1). P.B.M and S.E.P. acknowledge support from the National Institute for Health Research (NIHR) Barts Biomedical Research Centre. S.E.P. acknowledges support from the 'SmartHeart' EPSRC programme grant (www.nihr.ac.uk; EP/P001009/1) and from the CAP-AI programme, London's first AI enabling programme focused on stimulating growth in the capital's AI Sector. CAP-AI is led by Capital Enterprise in partnership with Barts Health NHS Trust and Digital Catapult and is funded by the European Regional Development Fund and Barts Charity. S.E.P. acts as a paid consultant to and is a shareholder of Circle Cardiovascular Imaging Inc., Calgary, Canada and Servier. SEP acknowledges the British Heart Foundation for funding the manual analysis to create a cardiovascular magnetic resonance imaging reference standard for the UK Biobank imaging resource in 5000 CMR scans (www.bhf.org.uk; PG/14/89/31194).PG, KL, and SEP have received funding from the European Union's 2020 research and innovation programme under grant agreement No 825903 (euCanSHare project). KL received funding from the Spanish Ministry of Science, Innovation and Universities under grant agreement RTI2018-099898-B-I00. NCH acknowledges support from the UK Medical Research Council (MRC #405050259, MRC LEU), NIHR Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton. N.A. recognize the National Institute for Health Research (NIHR) Integrated Academic Training programme which supports his Academic Clinical Lectureship posts. Z.R.E. was supported by British Heart Foundation Clinical Research Training Fellowship No. FS/17/81/33318. A.J. was supported by a Fulbright Predoctoral Research Award (2019-2020).

Abstract

Background: Cardiovascular magnetic resonance (CMR) radiomics analysis provides multiple quantifiers of ventricular shape and myocardial texture, which may be used for detailed cardiovascular phenotyping.

Objectives: We studied variation in CMR radiomics phenotypes by age and sex in healthy UK Biobank participants. Then, we examined independent associations of classical vascular risk factors (VRFs: smoking, diabetes, hypertension, high cholesterol) with CMR radiomics features, considering potential sex and age differential relationships.

Design: Image acquisition was with 1.5 Tesla scanners (MAGNETOM Aera, Siemens). Three regions of interest were segmented from short axis stack images using an automated pipeline: right ventricle, left ventricle, myocardium. We extracted 237 radiomics features from each study using Pyradiomics. In a healthy subset of participants (n=14,902) without cardiovascular disease or VRFs, we estimated independent associations of age and sex with each radiomics feature using linear regression models adjusted for body size. We then created a sample comprising individuals with at least one VRF matched to an equal number of healthy participants (n=27,400). We linearly modelled each radiomics feature against age, sex, body size, and all the VRFs. Bonferroni adjustment for multiple testing was applied to all p-values. To aid interpretation, we organised the results into six feature clusters.

Results: Amongst the healthy subset, men had larger ventricles with dimmer and less texturally complex myocardium than women. Increasing age was associated with smaller ventricles and greater variation in myocardial intensities. Broadly, all the VRFs were associated with dimmer, less varied signal intensities, greater uniformity of local intensity levels, and greater relative presence of low signal intensity areas within the myocardium. Diabetes and high cholesterol were also associated with smaller ventricular size, this association was of greater magnitude in men than women. The pattern of alteration of radiomics features with the VRFs was broadly consistent in men and women. However, the associations between intensity based radiomics features with both diabetes and hypertension were more prominent in women than men.

Conclusions: We demonstrate novel independent associations of sex, age, and major VRFs with CMR radiomics phenotypes. Further studies into the nature and clinical significance of these phenotypes are needed.

Keywords: cardiovascular magnetic resonance, radiomics, vascular risk factors, diabetes, hypertension, high cholesterol, smoking, sex, age.

Introduction

Epidemiologic studies highlight cigarette smoking, high blood pressure, and high cholesterol as major modifiable risk factors for cardiovascular disease(1,2). The association of these risk factors with incident cardiovascular events has been widely reported in multiple settings and their modification linked to substantial reductions in cardiovascular mortality(2).

There are important heterogeneities in cardiovascular disease patterns and clinical outcomes between men and women(3,4). These differences may be partly explained by differential biological consequences of vascular risk factors(5,6). Existing studies using cardiovascular magnetic resonance (CMR) have demonstrated distinct patterns of cardiovascular remodelling associated with classical vascular risk factors(7). Examining the potential sex differential impact of risk factors on cardiovascular phenotypes may provide insights into differences in cardiovascular disease patterns between men and women. However, this has not been addressed in existing work.

The application of radiomics analysis to CMR images allows extraction of multiple indices of ventricular shape and myocardial texture(8). Previous work has demonstrated the feasibility of CMR radiomics models for discrimination of health from disease(9–12), including distinction of vascular risk factors(13). These studies have focused on development of machine learning models optimised for disease discrimination using CMR radiomics features as input variables. CMR radiomics analysis may also be used for detailed cardiovascular phenotyping, with the potential to provide novel insights into disease processes. However, the approach of existing work does not allow granular evaluation of independent associations of CMR radiomics features with individual risk factors.

In this study, we demonstrate the utility of CMR radiomics analysis as a tool for detailed cardiovascular phenotyping. We characterise independent associations of sex, age, and key vascular risk factors with cardiovascular radiomics phenotypes and explore potential sex and age differential relationships.

Methods

Setting and study population

The UK Biobank is a very large cohort study comprising detailed characterisation of over 500,000 men and women from rural and urban settings across the UK. Individuals aged 40-69 years-old were identified from National Health Service (NHS) registers and recruited through postal invitation between 2006-2010. Individuals who were unable to consent or complete baseline assessment due to illness or discomfort were not included. There was baseline characterisation of demographics, lifestyle, and medical history of participants as well as blood sampling for selected biomarkers. The UK Biobank protocol is detailed in a dedicated document(14). The UK Biobank dataset is linked to routine national data sources including Hospital Episode Statistics (HES) and death registers, permitting continuous longitudinal tracking of incident health outcomes for the whole cohort(15). The UK Biobank imaging study, which includes, amongst other things, detailed CMR scanning, aims to image a random 20% (n=100,000) subset of the original participants. To date (June 2021), approximately 50,000 participants have completed the UK Biobank imaging study.

Background to CMR radiomics

The application of radiomics analysis to CMR images is a novel technique allowing extraction of quantitative measures of ventricular shape and myocardial texture. Image segmentations used for conventional image analysis may be used to define regions of interest for radiomics analysis, which typically include the ventricular cavities and the left ventricular (LV) myocardium. These segmentations are used to build 3D masks of the defined regions of interest, from which radiomics features are extracted. There are three categories of radiomics features: shape, first-order, and texture. The shape features provide advanced geometric quantification of the region of interest, including volume, axial dimensions, and quantitative descriptions of the overall shape (e.g., elongation, sphericity, flatness). The first-order and texture features are derived from analysis of the distribution and pattern of voxel signal intensity levels in the defined region of interest. The signal intensities in magnetic resonance images reflect magnetic properties of the underlying tissue, which are in turn influenced by tissue composition(16). Thus, radiomics signal intensity features applied to the LV myocardium may provide insight into myocardial tissue characteristics. First-order radiomics features describe the global distribution of signal intensities in the region of interest using histogram based statistics such as mean, variation, and skewness. Texture features rely on higher order statistics to describe local signal intensity patterns. Further details on CMR radiomics are provided in a dedicated review paper(8).

CMR image acquisition

The UK Biobank imaging study is performed using uniform pre-defined standard operating procedures, equipment, and staff training(17). CMR imaging was performed with 1.5 Tesla scanners (MAGNETOM Aera, Syngo Platform VD13A, Siemens Healthcare, Erlangen, Germany), the acquisition protocol is published elsewhere(18). Cardiac function assessment comprised three long axis cines (horizontal long axis, vertical long axis, left ventricular outflow tract sagittal and coronal) and a complete short axis stack covering the left and right ventricles acquired at one slice per breath hold using balanced steady-state free precession (bSSFP) sequences. Typical acquisition parameters are as follows: TR/TE = 2.6.1.1 ms, flip angle 80°, Grappa factor 2, voxel size 1.8 mm × 1.8 mm × 8 mm (6 mm for long axis). The actual temporal resolution of 32 ms was interpolated to 50 phases per cardiac cycle (~20 ms)(18). With the exception of distortion correction, no signal or image filtering was applied.

CMR image segmentation

The first 5,000 UK Biobank CMR scans were manually segmented using CVI[42]® post-processing software (Version 5.1.1, Circle Cardiovascular Imaging Inc., Calgary, Canada). The analysis protocol has been previously published(19). In brief, LV endocardial and epicardial borders were contoured in end-diastole and end-systole in the short axis stack images. End-diastole was defined as the first phase of the acquisition. End-systole was selected as the cardiac phase at which the mid-ventricular LV intra-cavity blood pool appeared smallest by visual inspection. The LV papillary muscles were considered part of the blood pool (excluded from LV mass). The right ventricular (RV) endocardial borders were segmented in end-diastole and end-systole. The most basal slice for the LV was included in the segmentation if at least half of the LV blood pool circumference was surrounded by myocardium. The pulmonary valve plane was used to define the most basal RV slice, with volumes below the valve plane considered as part of the RV. This ground truth manual analysis set, was used to develop a fully automated image analysis pipeline with inbuilt quality control(20). Details of reproducibility performance of the automated algorithm are available in dedicated publications(19–21). This pipeline has been propagated to the first 32,068 UK Biobank CMR studies, which, along with their corresponding segmentations, were available for inclusion in the present study.

Radiomics feature extraction

The segmentations from the short axis stack, described above, were used to define three regions of interest for radiomics analysis: RV cavity, LV cavity, LV myocardium. Features are calculated from 3D volumes of these ROIs. To reduce intensity level variations attributable to the acquisition process, we performed intensity normalisation of images through histogram matching, using as reference one of the studies from the dataset(22). For grey level discretisation, we used a fixed bin width of 25 intensity values. We extracted shape features from the RV and LV cavity. From the LV myocardium, we extracted signal intensity-based radiomics features (first order, texture). Radiomics features were extracted using the PyRadiomics open source platform version 2.2.0(23). Thus, a total of 237 radiomics features were included in the analysis for each CMR study (LV shape n= 26, RV shape n=26, LV myocardium first-order n=36, LV myocardium texture n=148). The full list of radiomics features included in the analysis is presented in Supplementary Table 1.

Feature clustering

As the number of radiomics was large, to aid interpretation, we grouped inter-correlated radiomics features using hierarchical cluster analysis (Figure 1)(24). More precisely, features were clustered using Ward's algorithm (Ward.D linkage function in R) so that variance is minimized within clusters with distance measured via Pearson coefficient (1-r)(25). The clusters were defined using features derived from participants free from cardiovascular disease and vascular risk factors. The optimal number of clusters was selected via consensus clustering using the ConsensusClusterPlus v1.50 function in R which allows for calculating quantitative stability evidence for determining the number and membership of possible clusters in an unsupervised manner(26). We assessed the curve for the change in the area under the Consensus Cumulative Distribution Function (CDF) and chose the number of clusters at which the area under the CDF no longer appreciably increases (the elbow). At six clusters, the CDF curve levelled off and all but one cluster had high consensus (Table 1, Figure 1), so we chose six clusters. We then assigned descriptive names to each cluster based on the properties of its constituent features, as summarised in Table 1.

Additionally, we examined correlation of conventional CMR metrics with all the radiomics features (Figure 1, Panel B). Conventional metrics correlated most strongly with radiomics features in the "size" cluster; correlation with other radiomics features was weak and inconsistent. Indicating that although there is some overlap between CMR radiomics features and conventional metrics, there are also many areas where radiomics features provide information that is different and uncorrelated to conventional metrics. Notably, LV mass additionally showed significant correlations with features in the "local variance" and "global uniformity" clusters. This may reflect dependency of these signal intensity-based features on ROI size (LV mass reflects the size of the myocardium ROI from which the texture features are extracted). It is also possible, that these metrics represent myocardial tissue alterations present in individuals with elevated LV mass (e.g. myocardial fibrosis).

Definition of the study sample

We first considered variation in radiomics features by sex and age in a healthy subset of participants. This analysis included participants without cardiovascular disease or vascular risk factors at time of imaging. For analysis of associations with vascular risk factors, we considered individuals who had vascular risk factors, but not cardiovascular disease. To create a balanced analysis sample, individuals with at least one vascular risk factor were matched on age and sex with participants without vascular risk factors (Supplementary Figure 1).

We considered cardiovascular disease as any ischaemic heart disease, non-ischaemic cardiomyopathy, valvular disease, or significant arrhythmia. These were ascertained from a combination of self-reported answers at baseline interview, UK Biobank algorithmically derived outcomes, and linked HES data codes (Supplementary Table 2). The following vascular risk factors were considered: hypertension, diabetes, high cholesterol, and current smoking. These were also defined by reference to a combination of self-reported answers, HES records, and blood biochemistry data (Supplementary Table 3). Age was taken as recorded at the time of imaging. Sex was taken from self-report at baseline.

Statistical analysis

Statistical analysis was performed using R version 3.6.222(27). Within the healthy subset, we estimated the independent associations of sex and age with individual radiomics features using multivariable linear regression models adjusted for body surface area. We calculated standardised beta coefficients, 95% confidence intervals, and p-values associated with age and sex for each radiomics feature. For ease of interpretation, we grouped these results within the previously defined feature clusters (Table 1). We calculated the average beta coefficient and confidence intervals for associations in each cluster. The full detail of associations of age and sex with individual radiomics features is presented in Supplementary Table 4.

To examine the association of vascular risk factors with radiomics features, we created a balanced cohort comprising a 1:1 ratio of "risk factor" and "no risk factor" individuals. To accomplish this, we estimated propensity scores from a logistic glm predicting presence of at least one risk factor from age and sex. Subjects with at least one risk factor were paired with their nearest neighbour with no risk factor using the R package matchit 4.1.0(28). Thus, the analysis sample comprised an equal number of individuals with vascular risk factors and those without vascular risk factors matched on age and sex. Within this sample, we entered all the vascular risk factors in a mutually adjusted multivariable linear regression model to estimate the independent association of each risk factor with individual radiomics features adjusting for age, sex, and body surface area. As before, we organise these results within the

previously defined clusters, reporting the average beta coefficient and confidence interval for each cluster. We present the results for associations of each vascular risk factor with individual radiomics features in Supplementary Table 5.

For all associations, we tested for potential differential relationships by sex and age, using interaction terms in fully adjusted models and explored the nature of any significant interactions in stratified analyses. We adjusted for multiple testing using a conservative Bonferroni correction per number of features (p*237).

Results

Baseline participant characteristics

CMR data was available for 32,068 UK Biobank participants, comprising 15,443 (48.2%) men and 16,625 women (51.8%) with average age of 63.3 ± 7.5 years (Table 2). The rates of diabetes, high cholesterol, hypertension, and smoking were 5.9%, 34.8%, 32.9%, and 3.6% respectively (Table 2). Ischaemic heart disease was the most common cardiovascular disease and was observed in 6.0% of participants (Table 2). Overall, there were 3,528 (11.0%) participants with documented cardiovascular disease (Supplementary Figure 1).

Exclusion of individuals with cardiovascular disease and vascular risk factors, resulted in a sample of 14,902 participants, which were considered as the healthy subset. This cohort comprised 6,095 men and 8,807 women, with mean ages of 61.5 ± 7.6 years and 60.7 ± 7.1 years, respectively (Table 2). The matched cohort comprised 13,700 individuals with at least one vascular risk factor matched 1:1 on age and sex to healthy participants creating a total analysis sample of 27,400 participants (Supplementary Figure 1, Table 2)

Variation of radiomics features by age and sex in the healthy subset

Associations of sex with radiomics features in the healthy subset

We estimated the association of sex with radiomics features in the healthy subset, whilst adjusting for age and body size. Full details of all linear regression coefficients and p-values are presented in Supplementary Table 4. For ease of interpretation, we group associations into previously defined feature clusters and calculate the mean beta coefficient for each cluster (Table 3, Figure 2).

There were significant associations between sex and radiomics features across all feature clusters. Compared to women, men had larger ventricular cavity sizes ("size" cluster, average beta: 0.58, 95% CI: 0.51, 0.66), with a less spherical overall shape of the ventricles ("shape" cluster, mean beta: -0.28, 95% CI: -0.36, -0.19), these shape alterations were broadly consistent for the LV and RV (Supplementary Table 4). There were also distinct differences in the distribution and patterns of signal intensities of the LV myocardium for men and women. Men had, on average, lower global signal intensity values ("global intensity" cluster, mean beta: -0.24, 95% CI: -0.33, -0.16) and less variation in intensity values ("global variance" cluster, average beta: -0.90, 95% CI: -0.97, -0.84). Furthermore, men showed enhanced measures of local dimness patterns ("local dimness" cluster, mean beta: 0.19, 95% CI: 0.02, 0.36) indicating greater relative presence of areas of low signal intensity in the LV myocardium compared to women. Consistent with this observation, men also had greater local uniformity of myocardial signal intensities ("local Uniformity" cluster, mean beta: 0.76, 95% CI: 0.68, 0.84), indicating a more homogeneous appearance of myocardial signal intensity levels. Thus, overall, compared to women men had larger more elongated ventricles with dimmer and less texturally complex appearance of the LV myocardium intensities.

Associations of age with radiomics features in the healthy subset

We next considered, the association of age with each radiomics feature whilst adjusting for sex and body size. We report all linear modelling results in Supplementary Table 4. For ease of interpretation, we group associations into previously defined feature clusters and calculate the mean beta coefficient for each cluster (Table 3, Figure 2). Compared to associations between sex and radiomics features, there were fewer statistically significant associations with age and, in general, the magnitudes of effects were smaller.

As expected, older age was associated with smaller ventricular cavity size ("size" cluster, average beta: -0.12, 95% CI: -0.14, -0.10). The were no significant alterations of the overall ventricular shape with aging based on the mean associations within the shape cluster (beta: 0.02, 95% CI: -0.00, 0.05). Examination of individual feature associations revealed association of increasing age with less spherical LV and more spherical RV shape (Supplementary Table 4).

Older age was associated with greater variation in myocardial intensity levels ("global variance" cluster, mean beta: 0.07, 95% CI: 0.06, 0.09), but without significant alteration in the average myocardial brightness ("global intensity" cluster, mean beta: 0.02 95% CI: -0.00, 0.05). Corresponding to the increased variance, average local uniformity in textures decreased with increasing age ("local uniformity" cluster, mean beta: -0.05, 95% CI: -0.07, -0.03) and there was decrease in local dimness patterns ("local dimness" cluster, average beta: -0.02, 95% CI: -0.05, -0.00). Overall, myocardial signal intensity alterations with age appear mixed with a broad pattern indicating dimmer hearts in end systole and brighter hearts in end diastole.

Sex differential age-related alterations in radiomics features

We tested for potential sex differential age related alterations of radiomics features through consideration of interaction terms (sex*age) in models additionally adjusted for age, sex, and body size (Supplementary Table 4, Supplementary Figure 2, Table 3). Overall, aging related changes in radiomics features appeared consistent for men and women. Relatively few features show a significant sex-age interaction (n=55, 23%) and most clusters had a mean interaction effect close to zero (Supplementary Table 4, Supplementary Figure 2).

To further visualize variation of radiomics features with age in men and women, we plotted the mean z-scored radiomics value within each cluster stratified by sex across all ages (Figure 3). Overall, age-related changes in radiomics feature clusters were, on average, consistent for men and women. The local uniformity cluster had the largest number of features with statistically significant age-sex interactions (n=22). On average, men had higher local uniformity, which declined with age. Women had lower local uniformity compared to men with little change in the features within this cluster with aging.

Variation of radiomics features with vascular risk factors

In the matched cohort (n=27,400), we estimated the independent association of vascular risk factors with radiomics features in multivariable linear regression models mutually adjusted for all the risk factors and additionally adjusting for age, sex, and body surface area. Modelling results for the associations of the vascular risk factors with each radiomics feature are reported in Supplementary Table 5. For ease of interpretation, we group associations into previously defined feature clusters and calculate the mean beta coefficient for each cluster (Table 4, Figure 4). We discuss associations with each vascular risk factor in turn.

Associations of diabetes with radiomics features

The most prominent diabetes related alterations of radiomics features were within the size and global intensity clusters, with statistically significant associations in 93% (n=40) and 81% (n=42) of features within these clusters respectively. Diabetes was associated with decreased size of the LV and RV cavities ("size" cluster, mean beta: -0.20, 95% CI: -0.23, -0.17), decreased global intensity ("global intensity" cluster, mean beta: -0.17, 95% CI: -0.20, -0.14), lower global variance ("global variance" cluster, mean beta: -0.06, 95% CI: -0.07, -

0.04), and greater local dimness ("local dimness" cluster, mean beta: 0.05, 95% CI: 0.02, 0.08).

Associations at the mean were not significant for the local uniformity and shape clusters. However, considering the clusters more closely (Figure 4), we see that diabetes drives a differential response with both local uniformity and shape clusters. Since there is some within cluster heterogeneity in what features quantify, we examined the coefficient of individual features within each cluster (Supplementary Table 5). For example, within the shape cluster, a number of features quantify intensity variance, and these features trend downward (Supplementary Table 5). This corresponds well with the observed small but significant trend in global variance. Examination of individual feature associations reveals less spherical LV in end-diastole and more elongated RV in both end-diastole and end-systole (Supplementary Table 5). Overall, diabetes was associated with decreased ventricular size, decreased myocardial intensity (brightness), decreased global variance (variation in intensity levels), and increased local uniformity.

Sex and age differential associations of diabetes with radiomics features

To examine the potential sex and age differential association of diabetes with radiomics features, we first considered the separately computed interaction terms (Supplementary Table 6, Supplementary Figure 3). There was no evidence of an age differential relationship, with no significant interaction terms detected for any radiomics feature. For the most part, associations were also consistent for men and women, with a statistically significant interaction term observed in only 10% of radiomics features, the majority of these were from the size cluster (Table 4).

To inspect further, we separated the beta boxplots by sex and compared the distributions of diabetes associations for each cluster (Supplementary Figure 4). We found that no feature showed a difference in direction of average association. For size specifically, women showed a lower average effect size than for men.

Associations of high cholesterol with radiomics features

High cholesterol had a unique signature of radiomic changes (Table 4, Figure 4). Like diabetes, high cholesterol was associated with smaller ventricular size ("size" cluster, mean beta: -0.09, 95% CI: -0.10, 0.08), however the magnitude of this association was smaller than that for diabetes and was not statistically significant. Examination of individual features within the "shape" cluster (specifically: sphericity, elongation, flatness), revealed differential shape associations in the LV and RV, with less sphericity of the former and greater sphericity of the latter (Supplementary Table 5). High cholesterol was also associated with decreased global intensity and slightly increased local dimness. Like diabetes, high cholesterol drives differential changes within the local uniformity cluster. Broadly, high cholesterol was associated with smaller ventricles, dimmer myocardium, and lower variance in myocardial intensities.

Sex and age differential associations of high cholesterol with radiomics features

We considered the impact of sex and age on the high cholesterol radiomics associations (Supplementary Table 6, Supplementary Figure 3). We identified few significant interaction effects for sex and age, 24% and 3% respectively (Table 4). The majority of the significant sex interactions were with features within the local uniformity (n=21) and global variance (n-18) clusters (Table 4). We therefore explored sex differential relationships within these clusters (Supplementary Figure 4). For both clusters, the direction of associations was

consistent for men and women, however the degree of the association can differ between the sexes (Supplementary Figure 4). As with diabetes, women showed a slightly lower size decrease with high cholesterol compared to men.

Associations of hypertension with radiomics features

Like diabetes and high cholesterol, hypertension was associated with significant decreases in global intensity of the LV myocardium ("global intensity" cluster, average beta: -0.07 95% CI: -0.09, -0.04). Hypertension was also associated with decreased variation in intensity levels ("global variance", mean beta: -0.14, 95% CI: -0.15, -0.13), increased local dimness ("local dimness, average beta: 0.07, 95% CI: 0.04, 0.10), and greater uniformity of local intensity levels ("local uniformity" cluster, average beta: 0.13, 95% CI: 0.11, 0.15). These myocardial alterations were the most consistent relationships observed with hypertension (Table 4, Figure 4).

For both the shape and size feature clusters, the significant associations appeared at the extremes of the beta coefficient distributions within each cluster, rather than at the mean (Figure 4). With regards the shape feature cluster, hypertension was associated with more elongated, less spherical ventricular shapes based on the average cluster association ("shape" cluster, average beta: -0.04, 95% CI: -0.06, -0.01). Examining individual feature associations, these associations appeared significant for the LV, but not the RV (Supplementary Table 5). The average beta coefficient in the size cluster demonstrated no significant association with hypertension. However, there were significant associations with a number of features (n=23) within this cluster, which lie distal either side of the distribution (Table 4, Figure 4).

Sex and age differential associations of hypertension with radiomics features

We examined potential variation of the associations of hypertension with radiomics features by sex and age (Supplementary Table 6, Supplementary Figure 3). The associations with hypertension were largely consistent across age and for men and women. There were significant interaction terms for sex and age in 23% and 7% of features respectively. Most of the features with significant sex interaction terms belonged to the global variance cluster (Table 4, Figure 4). In stratified analysis, we demonstrate that for both men and women, hypertension is associated with lower global variance; however, women show a greater decrease in global variance than men (Supplementary Figure 4).

Associations of smoking with radiomics features

Unlike the three previously considered vascular risk factors, smoking showed little consistent effect on any of the clusters of radiomics features (Table 4, Figure 4). The mean effect within each cluster is near zero (Figure 4). However, individual features show definite dependence on smoking (Supplementary Table 5). For example, end systolic global intensity features (e.g., mean and median signal intensities) all decreased with smoking. Furthermore, there were significant associations with RV shape features, demonstrating association of smoking with less spherical, flatter, and more elongated RV in both end-diastole and end-systole. These shape associations were not statistically significant with the LV (Supplementary Table 5).

In general, signal intensity based associations with smoking trended in similar directions to the other vascular risk factors. Broadly, the myocardium of smokers tends to decrease in global intensity and increase in local uniformity. However, these relationships were not as prominent as those for the other risk factors.

Sex and age differential associations of smoking with radiomics features

We found no evidence of differential associations of smoking with radiomics features by sex or age (Table 4, Supplementary Figure 3).

Discussion

Summary of findings

In this large study of UK Biobank participants free from cardiovascular disease, we report novel independent associations of CMR radiomics features with sex, age, diabetes, high cholesterol, hypertension, and smoking.

Amongst healthy participants, whilst adjusting for sex and body size, men had larger more elongated ventricles with dimmer, more homogenous, and less texturally complex appearance of the myocardium compared to women. In healthy aging, we observed smaller ventricular sizes and greater variation in myocardial signal intensity levels with increasing age, independent of sex and body size.

The pattern of associations with myocardial signal intensity features were broadly similar across vascular risk factors; all were associated with dimmer less varied myocardial signal intensities, greater uniformity of local intensity levels, and greater relative presence of low signal intensity areas. These independent associations with signal intensity phenotypes appeared most prominent with first hypertension and second diabetes. Both diabetes and high cholesterol were associated with smaller ventricular sizes, which appeared of greater magnitude for diabetes. Hypertension was associated with an overall less spherical, more elongated LV shape. Associations with smoking were of smaller magnitude than with other risk factors. Broadly, smoking was associated with significant alteration of RV, but not LV shape features.

In general, these relationships appeared consistent for men and women and across ages. Trends with healthy aging appeared consistent for men and women, and sex interactions, generally, indicated greater rapidity of age-related phenotypic alterations in men. The associations of diabetes with smaller ventricular size were a prominent feature for diabetic men, but not for women, in whom myocardial intensity features dominated. The association of hypertension with myocardial signal intensity phenotypes also varied by sex with hypertensive women showing a greater decrease in global variance than men.

Comparison with existing work

Our findings of larger ventricular sizes in healthy men compared to women (after adjustment for body size) and reduced ventricular size in healthy aging are consistent with previous studies using conventional CMR measures(29,30). Our additional observations relating to greater elongation of male hearts as well as myocardial signal intensity variations have not been previously described. Notably the differences in signal intensity patterns of male hearts resemble alterations we observed in association with vascular risk factors. That is, both male sex and vascular risk factors were associated with dimmer myocardial signal intensities, less variation in intensity patterns, and a more homogeneous appearance of the myocardium. This indicates that, in general, adverse cardiovascular exposures have some common manifestations in radiomics myocardial signal intensity features, perhaps indicating a shared pathophysiological process. Indeed, in a previous study of the associations between meat intake and cardiovascular phenotypes, we observed association of greater red and processed meat intake (adverse exposures) with dimmer and less varied myocardial signal intensities(31). The observation of these same phenotypes in healthy men suggests either undiagnosed vascular risk factors in men, or generally a poorer exposure profile in men than women with regards non-classical risk factors.

The cardiovascular phenotyping of vascular risk factors using conventional analysis of non-invasive imaging has been widely described. Our findings of smaller ventricular sizes associated with diabetes and high cholesterol are consistent with previous studies of the UK Biobank and the Multi-ethnic Study of Atherosclerosis (MESA) cohorts, using conventional CMR analysis(7,32). In addition, we demonstrate association of male sex and hypertension with alteration of the overall ventricular geometry towards a more elongated shape.

Myocardial intensity alterations were a prominent phenotype of diabetes and hypertension in our study, indicating that myocardial level alterations are key features of these conditions. Previous studies using echocardiography have demonstrated alteration of myocardial acoustic properties, an indicator of myocardial fibrosis, in diabetes and the correlation of this feature with diabetic disease severity and associated complications(33,34). Similarly, CMR studies using global contrast enhanced myocardial T1 mapping methods, have demonstrated that greater myocardial fibrosis (shorter T1 on contrast enhanced T1 mapping) in patients with diabetes is associated with poorer global longitudinal strain and diastolic dysfunction(35). There are also multiple reports of myocardial scarring and diffuse fibrosis associated with hypertension detectable using contrast and non-parametric mapping CMR techniques(36–39). Thus, it appears likely that myocardial fibrosis is a key component of the pathophysiology of both diabetic and hypertensive cardiomyopathies and that this may be detected using non-invasive imaging. The myocardial intensity alterations in our results also extended to high cholesterol, male sex, and (to a lesser extent) smoking. In a large study of the MESA cohort, Turkbey et al.(37) report associations of male sex, hypertension, and smoking with myocardial fibrosis detected by late gadolinium enhancement CMR images. The myocardial signal intensities in magnetic resonance imaging reflect the magnetic properties of underlying tissue, which in turn are determined by tissue characteristics(16). Thus, it is likely that our observations of signal intensity alterations reflect myocardial tissue characteristics, considered in the context of previous work, these may indicate diffuse myocardial fibrosis as a common pathophysiological process for the conditions considered.

Overall, the patterns of associations were consistent for men and women. There was evidence of potential sex differential alterations for selected features in diabetes and hypertension. In general, myocardial intensity alterations appeared a more important manifestation of these conditions in women than men, possibly indicate greater myocardial fibrosis in women. This observation is consistent with clinical observations of greater propensity for heart failure and specifically heart failure preserved ejection fraction syndromes in women, particularly in the context of diabetes and hypertension(40–43).

In summary, our findings with CMR radiomics analysis support previous reports using echocardiography and conventional CMR and provide more granular quantification of myocardial alterations and novel shape features associated with classical vascular risk factors in a low-risk group without clinically manifest cardiovascular disease.

Technical considerations

We adopted several technical approaches for increasing the clarity and statistical power of our results, but these approaches come with assumptions and limitations. First, to derive interpretable groups of related radiomics features, we clustered the features by their correlation in the healthy cohort. In doing this, we assumed that the healthy human population provided the best baseline to define the relationship between radiomics features. However, this approach skewed our identified clusters to group features that naturally correlate in human populations rather than features that correlate *definitionally*. For example,

myocardial intensity variance in end systole is in the Global Intensity cluster while myocardial intensity variance in end diastole is in the Global Variance cluster. If we had derived our clusters from digital phantoms instead (44), these two measures of intensity variance would have clustered together. We ultimately argue that clustering by human data works well for interpretability but encourage future studies to consider clustering on phantoms for better "ground truth" associations, although this may not always be feasible.

Another assumption of our work is that controlling for a linear association with BSA is sufficient to control for the relationship between radiomics features and body size. The confounding association between radiomics features and ROI size is well known (45,46), and we accounted for this by adjusting our linear regression for participants' BSA. However, it is also likely that radiomics features have complex nonlinear relationships with BSA. Therefore, a set of adjustments with nonlinear BSA terms in our linear modelling could produce better controls for BSA. However, an optimal approach to body size adjustment of radiomics features is yet to be established and adjustment for BSA in the context of the present study was deemed adequate.

Strengths and limitations

The large well characterised cohort in this study permitted reliable ascertainment of diseases and risk factors of interest. CMR image acquisition and segmentation was performed uniformly for the whole dataset minimising related technical variations. We demonstrate the feasibility of CMR radiomics and its application as a tool for deep cardiovascular phenotyping. Whilst previous studies do not consider confounding, we present associations adjusted for all vascular risk factors, body size, age, and sex. However, there may be other important confounds not considered here. This may be particularly relevant in understanding sex differences in associations, as we know that men and women differ in many other important ways not considered in our models. Associations of non-classical risk factors with radiomics phenotypes and their potential modifying effects on the relationships described in the present study is warranted. For instance, exploration of the influence of environmental, socio-demographic, and early life exposures on cardiac phenotypes may provide novel insights into the impact of these factors on cardiovascular health. The UK Biobank comprised a narrow age range, which may have limited our ability to detect age related alterations in CMR metrics. Exploration of age-related radiomics changes in a cohort with broader spectrum of ages is warranted. Furthermore, validation of our findings in different cohorts and within multi-centre settings is indicated in future work. A key avenue for future research is examining the correlation and incremental clinical value of CMR radiomics, particularly the signal intensity based features, against conventional measures of myocardial tissue character (e.g., native T1, late gadolinium enhancement). Due to the observational nature of the study, we cannot exclude residual confounding or infer causation (in either direction) from our results. Finally, there is need for dedicated studies to understand the biological and clinical significance of these radiomics phenotypes. Understanding the nature of these disease associations can be helpful for future studies with non-classical exposures, where the importance to cardiovascular health may not be so well understood. Additionally, investigating the incremental utility of radiomics analysis to predict incident health outcomes is a key research question in development of the technique as a novel imaging biomarker.

Conclusions

In this study we characterise novel associations of sex, age, and major vascular risk factors with cardiovascular radiomics phenotypes. These observations provide new insights into the impact of these risk factors on cardiovascular health, including potential sex differential

patterns of remodelling. Further studies into the nature and clinical significance of the defined phenotypes are needed.

Ethics

This study complies with the Declaration of Helsinki; the work was covered by the ethical approval for UK Biobank studies from the NHS National Research Ethics Service on 17th June 2011 (Ref 11/NW/0382) and extended on 18 June 2021 (Ref 21/NW/0157) with written informed consent obtained from all participants.

Author Contributions Statement

SEP, KL, NCH and ZRE conceived the idea. AJ led and conducted the analysis. PG extracted and prepared the radiomics features. CM contributed to data preparation. ZRE and AJ wrote the manuscript. All co-authors read and provided critical feedback on the manuscript.

Contribution to the Field Statement

Application of radiomics analysis to cardiovascular magnetic resonance (CMR) images is a novel analysis technique allowing extraction of multiple cardiac shape and myocardial texture features. This methodology may be used to derive deep cardiovascular phenotypes.

Previous work has demonstrated the feasibility of CMR radiomics analysis and the utility of radiomics features for cardiovascular disease discrimination, including distinction of vascular risk factors. However, these models are optimised for disease discrimination, rather than deep phenotyping. Furthermore, these studies do not consider sex differential patterns or account for confounding from co-existence of multiple risk factors.

In this large study of UK Biobank participants free from cardiovascular disease, we characterise novel associations of major vascular risk factors (sex, age, diabetes, high cholesterol, hypertension, and smoking) with cardiovascular radiomics phenotypes. These observations provide new insights into the impact of these risk factors on cardiovascular health, including potential sex differential patterns of remodelling.

Conflict of Interest Statement

SEP provides consultancy to and owns stock of Cardiovascular Imaging Inc, Calgary, Alberta, Canada.

References

1. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: A historical perspective. *Lancet* (2014) **383**:999–1008. doi:10.1016/S0140-6736(13)61752-3
2. Vilahur G, Badimon JJ, Bugiardini R, Badimon L. Perspectives: The burden of cardiovascular risk factors and coronary heart disease in Europe and worldwide. *Eur Hear Journal, Suppl* (2014) **16**:A7–A11. doi:10.1093/eurheartj/sut003
3. Haider A, Bengs S, Luu J, Osto E, Siller-Matula JM, Muka T, Gebhard C. Sex and gender in cardiovascular medicine: Presentation and outcomes of acute coronary syndrome. *Eur Heart J* (2020) **41**:1328–1336. doi:10.1093/eurheartj/ehz898
4. Townsend N, Wilson L, Bhatnagar P, Wickramasinghe K, Rayner M, Nichols M. Cardiovascular disease in Europe: Epidemiological update 2016. *Eur Heart J* (2016) **37**:3232–3245. doi:10.1093/eurheartj/ehw334
5. Gerdts E, Regitz-Zagrosek V. Sex differences in cardiometabolic disorders. *Nat Med* (2019) **25**:1657–1666. doi:10.1038/s41591-019-0643-8
6. Wang Y, O'Neil A, Jiao Y, Wang L, Huang J, Lan Y, Zhu Y, Yu C. Sex differences in the association between diabetes and risk of cardiovascular disease, cancer, and all-cause and cause-specific mortality: A systematic review and meta-analysis of 5,162,654 participants. *BMC Med* (2019) **17**: doi:10.1186/s12916-019-1355-0
7. Petersen SE, Sanghvi MM, Aung N, Cooper JA, Paiva JM, Zemrak F, Fung K, Lukaschuk E, Lee AM, Carapella V, et al. The impact of cardiovascular risk factors on cardiac structure and function: Insights from the UK Biobank imaging enhancement study. *PLoS One* (2017) **12**:45–52. doi:10.1371/journal.pone.0185114
8. Raisi-Estabragh Z, Izquierdo C, Campello VM, Martin-Isla C, Jaggi A, Harvey NC, Lekadir K, Petersen SE. Cardiac magnetic resonance radiomics: basic principles and clinical perspectives. *Eur Hear J - Cardiovasc Imaging* (2020) **21**:349–356. doi:10.1093/ehjci/jeaa028
9. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R. Radiomic Analysis of Myocardial Native T1 Imaging Discriminates Between Hypertensive Heart Disease and Hypertrophic Cardiomyopathy. *JACC Cardiovasc Imaging* (2019)1–9. doi:10.1016/j.jcmg.2018.11.024
10. Baessler B, Luecke C, Lurz J, Klingel K, von Roeder M, de Waha S, Besler C, Maintz D, Gutberlet M, Thiele H, et al. Cardiac MRI Texture Analysis of T1 and T2 Maps in Patients with Infarctlike Acute Myocarditis. *Radiology* (2018) **289**:357–365. doi:10.1148/radiol.2018180411
11. Baeßler B, Mannil M, Maintz D, Alkadhi H, Manka R. Texture analysis and machine learning of non-contrast T1-weighted MR images in patients with hypertrophic cardiomyopathy—Preliminary results. *Eur J Radiol* (2018) **102**:61–67. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0720048X18300962 [Accessed September 17, 2019]
12. Cheng S, Fang M, Cui C, Chen X, Yin G, Prasad SK, Dong D, Tian J, Zhao S. LGE-CMR-derived texture features reflect poor prognosis in hypertrophic cardiomyopathy patients with systolic dysfunction: preliminary results. *Eur Radiol* (2018) **28**:4615–4624. doi:10.1007/s00330-018-5391-5
13. Cetin I, Raisi-Estabragh Z, Petersen SE, Napel S, Piechnik SK, Neubauer S, Gonzalez Ballester MA, Camara O, Lekadir K. Radiomics Signatures of Cardiovascular Risk Factors in Cardiac MRI: Results From the UK Biobank. *Front Cardiovasc Med* (2020) **7**:591368. doi:10.3389/fcvm.2020.591368
14. UK Biobank Coordinating Centre. UK Biobank: Protocol for a large-scale prospective epidemiological resource. *UKBB-PROT-09-06 (Main Phase)* (2007) **06**:1–112.

Available at: https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf [Accessed December 13, 2019]
15. Raisi-Estabragh Z, Petersen SE. Cardiovascular research highlights from the UK Biobank: opportunities and challenges. *Cardiovasc Res* (2020) **116**:e12–e15. doi:10.1093/cvr/cvz294
16. Lombardi M, Plein S, Petersen S, Bucciarelli-Ducci C, Buechel EV, Basso C, Ferrari V eds. *The EACVI Textbook of Cardiovascular Magnetic Resonance*. Oxford University Press (2018). doi:10.1093/med/9780198779735.001.0001
17. Raisi-Estabragh Z, Harvey NC, Neubauer S, Petersen SE. Cardiovascular magnetic resonance imaging in the UK Biobank: a major international health research resource. *Eur Hear J - Cardiovasc Imaging* (2021) **22**:251–258. doi:10.1093/ehjci/jeaa297
18. Petersen SE, Matthews PM, Francis JM, Robson MD, Zemrak F, Boubertakh R, Young AA, Hudson S, Weale P, Garratt S, et al. UK Biobank's cardiovascular magnetic resonance protocol. *J Cardiovasc Magn Reson* (2016) **18**:8. doi:10.1186/s12968-016-0227-4
19. Petersen SE, Aung N, Sanghvi MM, Zemrak F, Fung K, Paiva JM, Francis JM, Khanji MY, Lukaschuk E, Lee AM, et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort. *J Cardiovasc Magn Reson* (2017) **19**:18. doi:10.1186/s12968-017-0327-9
20. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* (2018) **20**:65. doi:10.1186/s12968-018-0471-x
21. Attar R, Pereañez M, Gooya A, Albà X, Zhang L, de Vila MH, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, et al. Quantitative CMR population imaging on 20,000 subjects of the UK Biobank imaging study: LV/RV quantification pipeline and its evaluation. *Med Image Anal* (2019) **56**:26–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/31154149 [Accessed October 7, 2019]
22. Gonzalez R, Fittes B. 2nd Conference on Remotely Manned Systems: Technology and Applications. in *Gray-level transformations for interactive image enhancement* (Los Angeles, California), 17–19. Available at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19770022806.pdf
23. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) **77**:e104–e107. doi:10.1158/0008-5472.CAN-17-0339
24. Maechler M. "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al. R Packag. version 2.0. (2019) Available at: https://www.rdocumentation.org/packages/cluster/versions/2.1.0 [Accessed May 3, 2020]
25. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* (2014) **31**:274–295. doi:10.1007/s00357-014-9161-z
26. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach Learn* (2003) **52**:91–118. doi:10.1023/A:1023949509487
27. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/ [Accessed October 18, 2020]

28. Ho DE, Imai K, King G, Stuart EA. MatchIt : Nonparametric Preprocessing for Parametric Causal Inference. *J Stat Softw* (2011) **42**:1–28. doi:10.18637/jss.v042.i08
29. Raisi-Estabragh Z, Kenawy AAM, Aung N, Cooper J, Munroe PB, Harvey NC, Petersen SE, Khanji MY. Variation in left ventricular cardiac magnetic resonance normal reference ranges: systematic review and meta-analysis. *Eur Hear J - Cardiovasc Imaging* (2021) **22**:494–504. doi:10.1093/ehjci/jeaa089
30. Kawel-Boehm N, Hetzel SJ, Ambale-Venkatesh B, Captur G, Francois CJ, Jerosch-Herold M, Salerno M, Teague SD, Valsangiacomo-Buechel E, van der Geest RJ, et al. *Reference ranges ("normal values") for cardiovascular magnetic resonance (CMR) in adults and children: 2020 update*. BioMed Central (2020). doi:10.1186/s12968-020-00683-3
31. Raisi-Estabragh Z, McCracken C, Gkontra P, Jaggi A, Ardissino M, Cooper J, Biasiolli L, Aung N, Piechnik SK, Neubauer S, et al. Associations of Meat and Fish Consumption With Conventional and Radiomics Cardiovascular Magnetic Resonance Phenotypes in the UK Biobank. *Front Cardiovasc Med* (2021) **8**: doi:10.3389/fcvm.2021.667849
32. Liu CY, Lai S, Kawel-Boehm N, Chahal H, Ambale-Venkatesh B, Lima JAC, Bluemke DA. Healthy aging of the left ventricle in relationship to cardiovascular risk factors: The Multi-Ethnic Study of Atherosclerosis (MESA). *PLoS One* (2017) **12**: doi:10.1371/journal.pone.0179947
33. Pérez JE, McGill JB, Santiago J V., B. Schechtman K, Waggoner AD, Miller JG, Sobel BE. Abnormal myocardial acoustic properties in diabetic patients and their correlation with the severity of disease. *J Am Coll Cardiol* (1992) **19**:1154–1162. doi:10.1016/0735-1097(92)90316-F
34. Di Bello V, Talarico L, Picano E, Di Muro C, Landini L, Paterni M, Matteucci E, Giusti C, Giampietro O. Increased echodensity of myocardial wall in the diabetic heart: An ultrasound tissue characterization study. *J Am Coll Cardiol* (1995) **25**:1408–1415. doi:10.1016/0735-1097(95)00026-Z
35. Ng ACT, Auger D, Delgado V, Van Elderen SGC, Bertini M, Siebelink HM, Van Der Geest RJ, Bonetti C, Van Der Velde ET, De Roos A, et al. Association between diffuse myocardial fibrosis by cardiac magnetic resonance contrast-enhanced T1 mapping and subclinical myocardial dysfunction in diabetic patients a pilot study. *Circ Cardiovasc Imaging* (2012) **5**:51–59. doi:10.1161/CIRCIMAGING.111.965608
36. Mavrogeni S, Katsi V, Vartela V, Noutsias M, Markousis-Mavrogenis G, Kolovou G, Manolis A. The emerging role of Cardiovascular Magnetic Resonance in the evaluation of hypertensive heart disease. *BMC Cardiovasc Disord* (2017) **17**:1–10. doi:10.1186/s12872-017-0556-8
37. Turkbey EB, Nacif MS, Guo M, McClelland RL, Teixeira PBRP, Bild DE, Barr RG, Shea S, Post W, Burke G, et al. Prevalence and correlates of myocardial scar in a US cohort. *JAMA - J Am Med Assoc* (2015) **314**:1945–1954. doi:10.1001/jama.2015.14849
38. Rodrigues JCL, Amadu AM, Dastidar AG, Szantho G V., Lyen SM, Godsave C, Ratcliffe LEK, Burchell AE, Hart EC, Hamilton MCK, et al. Comprehensive characterisation of hypertensive heart disease left ventricular phenotypes. *Heart* (2016) **102**:1671–1679. doi:10.1136/heartjnl-2016-309576
39. Schumann CL, Jaeger NR, Kramer CM. Recent Advances in Imaging of Hypertensive Heart Disease. *Curr Hypertens Rep* (2019) **21**:3. doi:10.1007/s11906-019-0910-6
40. Scantlebury DC, Borlaug BA. Why are women more likely than men to develop heart failure with preserved ejection fraction? *Curr Opin Cardiol* (2011) **26**:562–568. doi:10.1097/HCO.0b013e32834b7faf

41. Ho JE, Gona P, Pencina MJ, Tu J V., Austin PC, Vasan RS, Kannel WB, D'Agostino RB, Lee DS, Levy D. Discriminating clinical features of heart failure with preserved vs. reduced ejection fraction in the community. *Eur Heart J* (2012) **33**:1734–1741. doi:10.1093/eurheartj/ehs070
42. Beale AL, Meyer PMD, Marwick TH, Lam CSP, Kaye DM. Sex differences in cardiovascular pathophysiology why women are overrepresented in heart failure with preserved ejection fraction. *Circulation* (2018) **138**:198–205. doi:10.1161/CIRCULATIONAHA.118.034271
43. Chadalavada S, Jensen MT, Aung N, Cooper J, Lekadir K, Munroe PB, Petersen SE. Women With Diabetes Are at Increased Relative Risk of Heart Failure Compared to Men: Insights From UK Biobank. *Front Cardiovasc Med* (2021) **8**: doi:10.3389/fcvm.2021.658726
44. Jaggi A, Mattonen SA, McNitt-Gray M, Napel S. Stanford DRO Toolkit: Digital Reference Objects for Standardization of Radiomic Features. *Tomography* (2020) **6**:111. doi:10.18383/J.TOM.2019.00030
45. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, Abdalah MA, Schabath MB, Goldgof DG, Mackin D, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys* (2017) **44**:1050–1062. doi:10.1002/mp.12123
46. Traverso A, Kazmierski M, Zhovannik I, Welch M, Wee L, Jaffray D, Dekker A, Hope A. Machine learning helps identifying volume-confounding effects in radiomics. *Phys Medica* (2020) **71**:24–30. doi:10.1016/J.EJMP.2020.02.010

Table 1. Summary of the six defined radiomics feature clusters including their assigned names, example features, and properties represented by the features within each cluster

Cluster Name	Example Features	Description of feature properties	Consensus D1
Size	Volume Surface Area	Size of the ventricles	0.98
Local Uniformity	First-order Uniformity GLSZM Large Area Emphasis	Size of areas with the same intensity level within myocardium	0.67
Global Variance	First-order Variance GLCM Contrast	Variance of myocardial intensity level distribution	0.51
Shape	Shape Elongation Shape Sphericity	Descriptors of overall ventricular shape	0.96
Local Dimness	GLDM Low Gray Level Emphasis GLSZM Low Gray Level Zone Emphasis	Relative presence of areas of low signal intensity level	0.78
Global Intensity	First-order Mean First-order Energy	Average brightness of myocardial intensity level	0.70

Table 1*. GLCM: Gray Level Co-occurrence Matrix; GLDM: Gray Level Dependence Matrix; GLSZM: Gray Level Size Zone Matrix. Consensus D1 indicates the repeatability of cluster components on repeated clustering, that is the likelihood that the same features appear in the cluster if the clustering analysis is repeated. Higher values within the shape category indicate greater sphericity and less elongated ventricular shapes. Please note, for computational reasons in Pyradiomics the "flatness" and "elongation" features are reported as inverse values, thus higher elongation and flatness values indicate less elongated more spherical shapes (and vice versa).*

Table 2. Baseline participant characteristics

	All participants	Healthy subset	Matched vascular risk factor cohort
Total Population	32068	14902	27400
Men	15443 (48.2%)	6095 (40.9%)	13290 (48.5%)
Women	16625 (51.8%)	8807 (59.1%)	14110 (51.5%)
Age at imaging (years)	63.3 ± 7.5	61.0 ± 7.3	63.4 ± 7.2
Body surface area (m^2)	1.9 ±0.2	1.8 ± 0.2	1.9 ± 0.2
Body mass index (Kg/m^2)	26.6 ± 4.2	25.6 ± 3.8	26.6 ± 4.2
Ischaemic heart disease	1937 (6.0%)	0	0
Valvular heart disease	582 (1.8%)	0	0
Non-ischaemic cardiomyopathies	59 (0.2%)	0	0
Heart failure unspecified aetiology	191 (0.6%)	0	0
Cardiac arrhythmia	1443 (4.5%)	0	0
Diabetes	1881 (5.9%)	0	1471
High cholesterol	11161 (34.8%)	0	8848
Hypertension	10545 (32.9%)	0	8322
Smoking (current)	1157 (3.6%)	0	1038

Table 2. *Continuous variables are summarised as mean ± standard deviation and count variables as number of participants (percentage of total).*

Table 3. Relationship of sex and age with radiomics features in the healthy subset expressed as the average association within each of the six radiomics feature clusters

		Radiomics feature clusters						
Exposures		**Size**	**Local Uniformity**	**Global Variance**	**Shape**	**Local Dimness**	**Global Intensity**	**Totals**
Sex (Male)	Mean Beta	0.58	0.76	-0.90	-0.28	0.19	-0.24	
	95% CI	0.51, 0.66	0.68, 0.84	-0.97, -0.84	-0.36, -0.19	0.02, 0.36	-0.33, -0.16	
	Significant features, n (%)	41 (95%)	45 (100%)	37 (100%)	34 (87%)	14 (70%)	43 (83%)	214 (91%)
Age	Mean Beta	-0.12	-0.05	0.07	0.02	-0.02	0.02	
	95% CI	-0.14, -0.10	-0.07, -0.03	0.06, 0.09	-0.00, 0.05	-0.05, -0.00	-0.00, 0.05	
	Significant features, n (%)	42 (98%)	37 (82%)	29 (78%)	27 (69%)	13 (65%)	46 (89%)	194 (82%)
Sex*Age	Mean Beta	-0.01	-0.07	0.03	0.02	0.00	-0.01	
	Lower CI	-0.015, -0.00	-0.08, -0.06	0.01, 0.04	0.00, 0.03	-0.02, 0.03	-0.02, 0.00	
	Significant features, n (%)	3 (7%)	22 (49%)	11 (30%)	7 (18%)	4 (20%)	8 (15%)	55 (23%)
	Total features in cluster (n)	43	45	37	39	20	52	236

Table 3. *Results are the mean beta coefficient and 95% CI for associations of each exposure with the features within each cluster. Beta indicates standard deviation change in radiomics feature per 1 unit/standard deviation change in the exposure. Models are mutually adjusted for age and sex, and include additional adjustment for body surface area. The interaction term is from a separate fully adjusted model. For sex, the reference level is set as "female". "Significant features" indicates the number and percentage of features with a statistically significantly association within each cluster, based on a Bonferroni adjusted p-value. CI: confidence interval.*

Table 4. Relationship of vascular risk factors with radiomics features in the healthy subset expressed as the average association within each of the six radiomics feature clusters

		Radiomics feature clusters						
Exposure		**Size**	**Local Uniformity**	**Global Variance**	**Shape**	**Local Dimness**	**Global Intensity**	**Totals**
Diabetes	Mean Beta	-0.20	0.006	-0.06	-0.01	0.05	-0.17	
	95% CI	-0.23, -0.17	-0.039, 0.05	-0.07, -0.04	-0.05, 0.04	0.02, 0.08	-0.20, -0.14	
	Significant features, n (%)	40 (93%)	15 (33%)	6 (16%)	17 (44%)	5 (25%)	42 (81%)	125 (53%)
Diabetes*Sex	Mean	-0.13	-0.050	0.094	0.028	-0.028	0.019	
	95% CI	-0.15, -0.11	-0.08, -0.02	0.08, 0.11	-0.01, 0.06	-0.05, -0.01	-0.00, 0.04	
	Significant features, n (%)	14 (33%)	4 (9%)	0 (0%)	3 (8%)	0 (0%)	2 (4%)	23 (10%)
Diabetes*Age	Mean	0.01	-0.00	0.01	-0.00	0.01	0.00	
	95% CI	0.00, 0.01	-0.01, 0.00	0.00, 0.01	-0.01, 0.01	0.00, 0.01	-0.01, 0.01	
	Significant features, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
High cholesterol	Mean	-0.09	-0.00	-0.01	0.00	0.05	-0.08	
	95% CI	-0.10, 0.08	-0.02, 0.02	-0.02, -0.01	-0.02, 0.02	0.04, 0.06	-0.09, 0.07	
	Significant features, n (%)	40 (93%)	15 (33%)	3 (8%)	19 (49%)	12 (60%)	37 (71%)	126 (53%)
High cholesterol*Sex	Mean	-0.04	-0.06	0.08	0.02	-0.01	0.01	
	95% CI	-0.05, -0.02	-0.08, -0.05	0.07, 0.09	0.00, 0.04	-0.04, 0.01	-0.00, 0.02	
	Significant features, n (%)	10 (23%)	21 (47%)	18 (49%)	3 (8%)	0 (0%)	4 (8%)	56 (24%)
High cholesterol*Age	Mean	0.03	-0.00	0.01	0.01	-0.01	0.02	
	95% CI	0.02, 0.03	-0.01, 0.00	0.01, 0.02	0.00, 0.02	-0.01, -0.01	0.02, 0.03	
	Significant features, n (%)	7 (16%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	7 (3%)
Hypertension	Mean	-0.00	0.13	-0.14	-0.04	0.07	-0.07	
	95% CI	-0.02, 0.01	0.11, 0.15	-0.15, -0.13	-0.06, -0.01	0.04, 0.10	-0.09, -0.04	
	Significant features, n (%)	23 (54%)	40 (89%)	37 (100%)	18 (46%)	15 (75%)	43 (83%)	176 (75%)
Hypertension*Sex	Mean	-0.03	-0.03	0.11	0.03	0.04	0.02	
	95% CI	-0.05, -0.02	-0.05, -0.01	0.10, 0.13	0.01, 0.05	0.02, 0.07	0.01, 0.03	
	Significant features, n (%)	5 (12%)	9 (20%)	25 (68%)	7 (18%)	2 (10%)	5 (10%)	53 (23%)
Hypertension*Age	Mean	0.02	-0.03	0.03	0.01	-0.01	0.02	
	95% CI	0.01, 0.02	-0.03, -0.02	0.03, 0.04	-0.00, 0.01	-0.02, -0.01	0.01, 0.03	
	Significant features, n (%)	1 (2%)	2 (4%)	7 (19%)	0 (0%)	0 (0%)	6 (12%)	16 (7%)
Smoking	Mean	-0.03	0.06	-0.06	-0.04	0.00	-0.06	

		Radiomics feature clusters						
Exposure		**Size**	**Local Uniformity**	**Global Variance**	**Shape**	**Local Dimness**	**Global Intensity**	**Totals**
	95% CI	-0.05, -0.01	0.04, 0.08	-0.07, -0.05	-0.07, -0.01	-0.02, -0.03	-0.08, -0.03	
	Significant features, n (%)	6 (14%	14 (31%)	4 (11%)	12 (31%)	0 (0%)	12 (23%)	48 (20%)
Smoking*Sex	Mean	-0.03	-0.01	0.05	0.05	0.05	-0.02	
	95% CI	-0.04, -0.01	-0.03, 0.00	0.04, 0.06	0.03, 0.07	0.02, 0.08	-0.04, 0.00	
	Significant features, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Smoking*Age	Mean	-0.02	-0.00	-0.01	-0.01	0.04	-0.01	
	95% CI	-0.03, -0.01	-0.01, 0.01	-0.01, 0.00	-0.02, -0.00	0.04, 0.05	-0.02, -0.01	
	Significant features, n (%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Total	43	45	37	39	20	52	236

Table 4. *Results are the mean beta coefficient and 95% CI for associations of each exposure with the features within each cluster. Beta indicates standard deviation change in radiomics feature per 1 unit/standard deviation change in the exposure. Models are mutually adjusted for all the risk factors (diabetes, high cholesterol, hypertension, smoking) and include adjustment for age, sex, and body surface area. Interaction terms are from separate fully adjusted models, separately for age and sex. "Significant features" indicates the number and percentage of features with a statistically significantly association within each cluster, based on a Bonferroni adjusted p-value. CI: confidence interval*

Figure 1. Illustrating the clustering method and approach to defining the number of radiomics feature clusters for the radiomics features

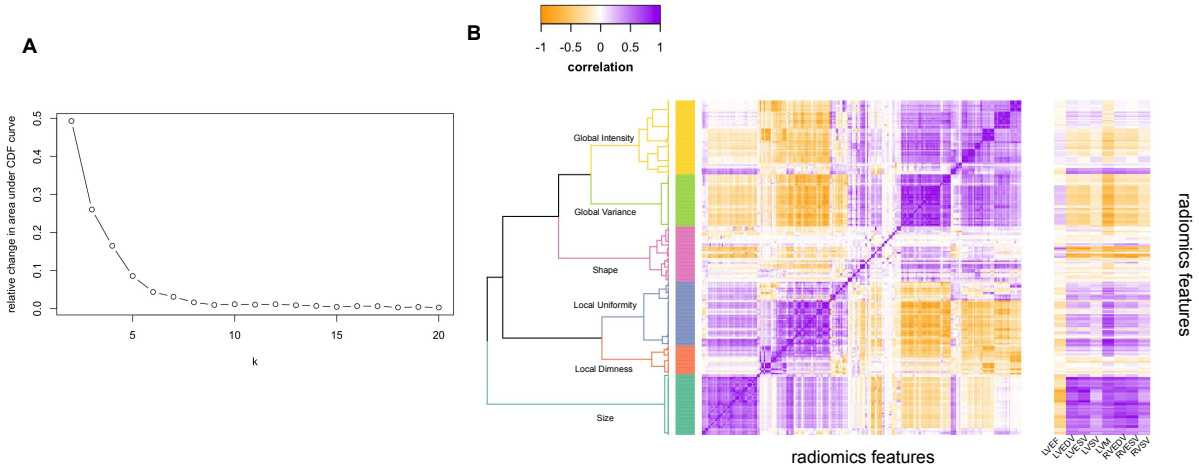

Figure 1*. Panel A illustrates the relative change in area under the CDF (Consensus Cumulative Distribution Function) curve of the y axis with increasing number of clusters (k on x axis), with the curve levelling off at six clusters. Panel B is the correlation heatmap illustrating the six defined clusters, with the darkest purple indicating perfect positive correlation and darkest yellow perfect negative correlation. The dendrogram indicates the six clusters from hierarchical clustering. The ribbon on the right of Panel B illustrates correlation of each radiomics feature with the conventional metrics indicated on the x-axis. LVEDV: left ventricular end-diastolic volume; LVEF: left ventricular ejection fraction; LVESV: left ventricular end-systolic volume; LVM: left ventricular mass; RVEDV: right ventricular end-diastolic volume; RVESV: right ventricular end-systolic volume; RVSV: right ventricular stroke volume.*

Figure 2. Associations of sex and age with radiomics features in the healthy subset grouped into clusters

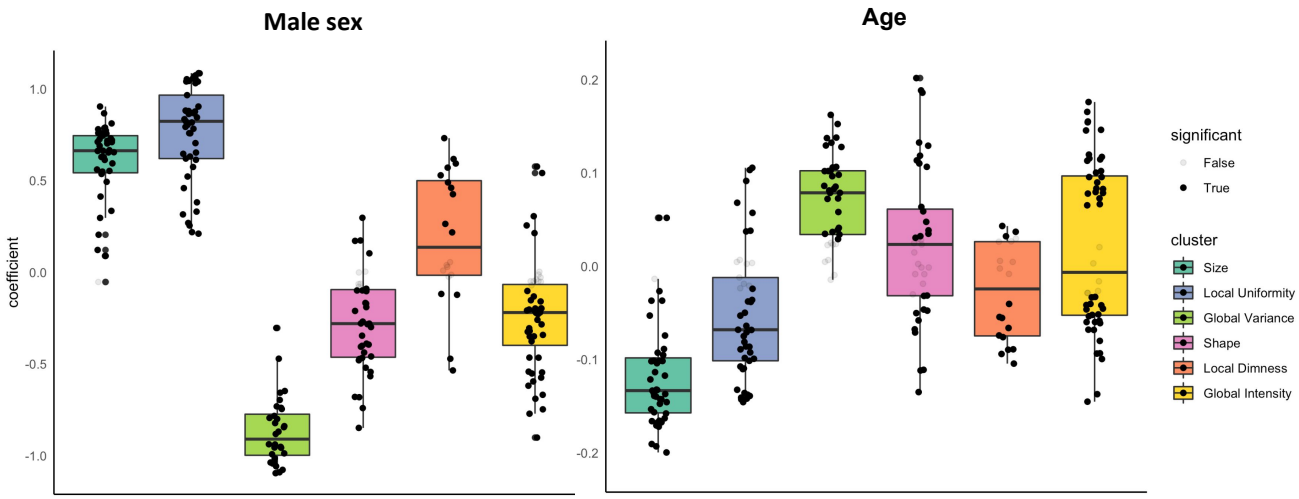

Figure 2*. Results are from linear regression models adjusted for age, sex, and body surface area. The y axis is standardised beta coefficients for associations of sex (left) and age (right) with radiomics features. Each dot represents point estimate of the association with a radiomic feature from a separate model. Black dots indicate statistically significant associations. Grey dots indicate non-significant associations. Statistical significance is based on Bonferroni adjusted p-value <0.05. Feature associations are grouped into previously defined clusters (Figure 1, Table 1). The dark line in the box plot indicates the median beta coefficient in the cluster, the box borders indicate limits of the interquartile range.*

Figure 3. Mean standardised radiomics value for each feature cluster stratified by sex across all ages

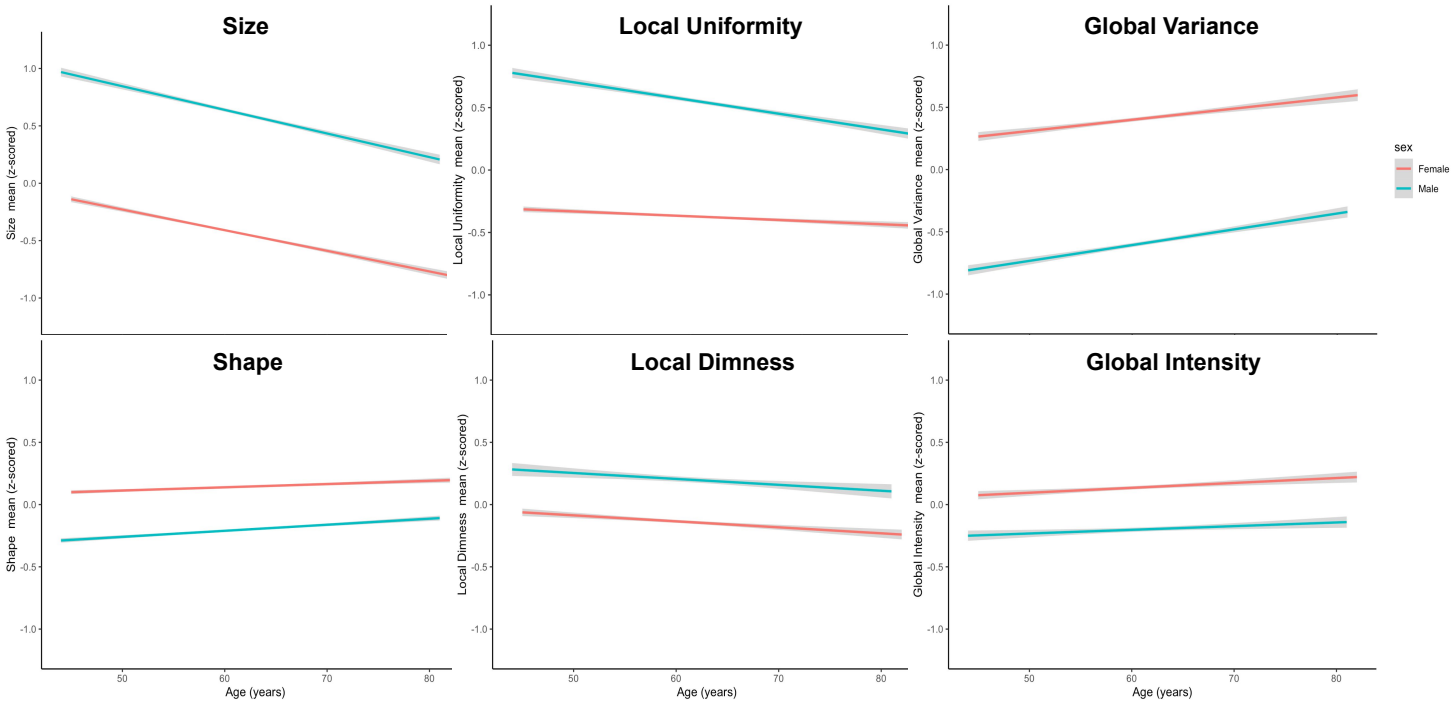

Figure 3. *Men had larger (higher size values) and more elongated (higher shape values) ventricles than women. Men had dimmer less varied signal intensities at both a global (lower global intensity, lower global variance) and local (higher local uniformity, higher local dimness) level. Alteration of radiomics features with aging were generally consistent for men and women. There was more rapid decline in local uniformity in men with minimal age-related change in this cluster for women.*

Figure 4. Associations of diabetes, high cholesterol, hypertension, and smoking with radiomics features grouped into clusters

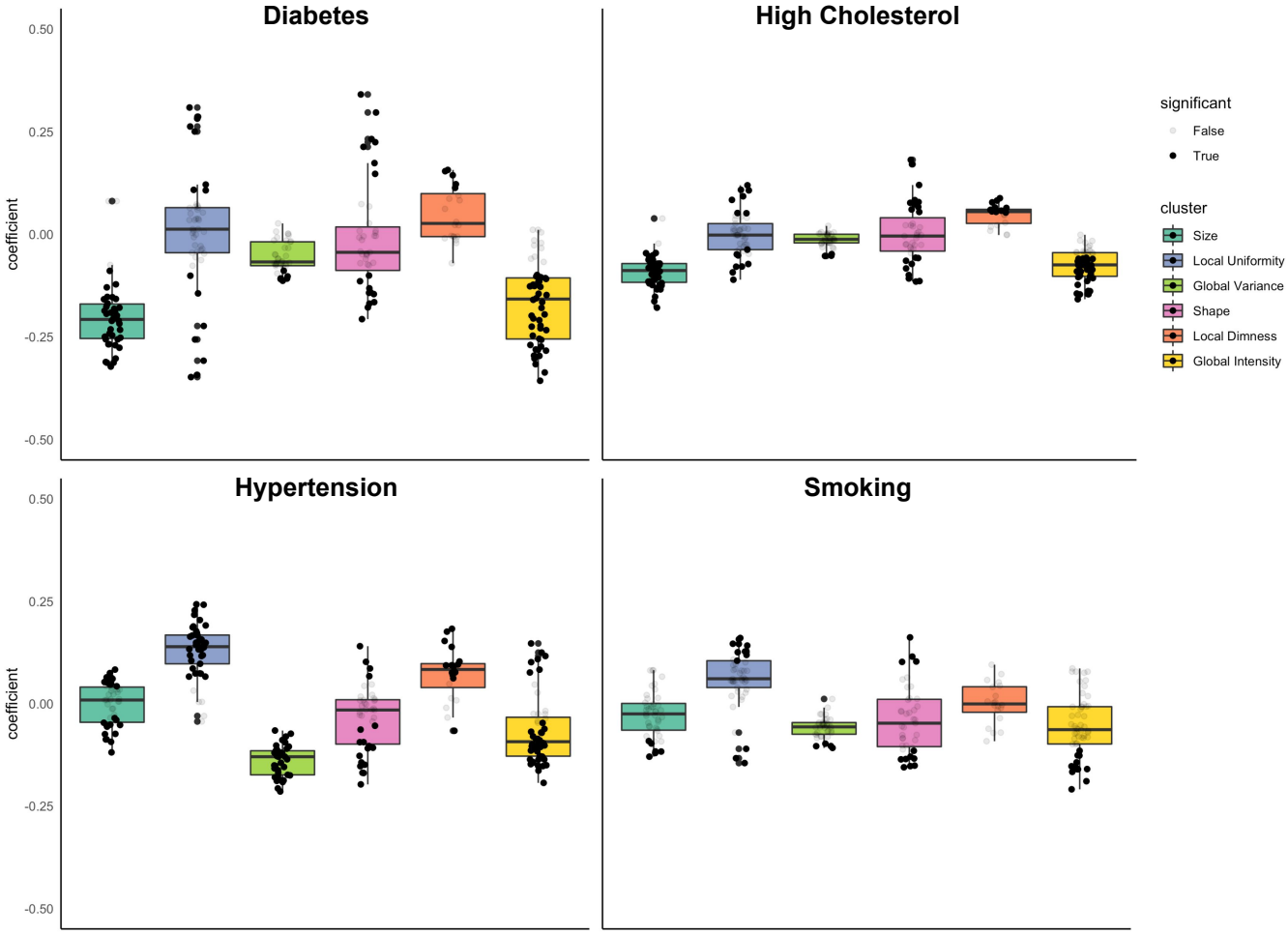

Figure 4. *Results are from linear regression models adjusted for age, sex, and body surface area, diabetes, high cholesterol, hypertension, and smoking. The y axis is standardised beta coefficients for associations of vascular risk factors (diabetes, high cholesterol, hypertension, smoking) with radiomics features. Each dot represents point estimate of association with a radiomic feature from a separate model. Black dots indicate statistically significant associations. Grey dots indicate non-significant associations. Statistical significance is based on Bonferroni adjusted p-value <0.05. Feature associations are grouped into previously defined clusters (Figure 1, Table 1). The dark line in the box plot indicates the median beta coefficient in the cluster, the box borders indicate limits of the interquartile range*