

A phenomapping-derived tool to personalize the selection of anatomical vs. functional testing in evaluating chest pain (ASSIST)

Evangelos K. Oikonomou ¹, David Van Dijk^{2,3}, Helen Parise ²,
Marc A. Suchard ^{4,5}, James de Lemos ⁶, Charalambos Antoniades⁷,
Eric J. Velazquez², Edward J. Miller ², and Rohan Khera ^{2,8*}

¹Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8056, USA; ²Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8056, USA; ³Department of Computer Science, Yale University, 51 Prospect St, New Haven, CT 06520-8285, USA; ⁴Department of Biostatistics, Fielding School of Public Health, University of California, 650 Charles E. Young Drive S, Los Angeles, CA 90095, USA; ⁵Departments of Computational Medicine and Human Genetics, David Geffen School of Medicine at UCLA, University of California, 695 Charles E. Young Drive S, Los Angeles, CA 90095, USA; ⁶Division of Cardiology, Department of Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-8830, USA; ⁷Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Level 6, West Wing, John Radcliffe Hospital, Headley Way, OX3 9DU, Oxford, UK; and ⁸Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, MS 1 Church Street, Suite 200, New Haven, CT 06510, USA

Received 17 February 2021; revised 14 March 2021; editorial decision 29 March 2021; accepted 31 March 2021; online publish-ahead-of-print 21 April 2021

See page 2549 for the editorial comment on this article (doi:10.1093/eurheartj/ehab313)

Aims

Coronary artery disease is frequently diagnosed following evaluation of stable chest pain with anatomical or functional testing. A more granular understanding of patient phenotypes that benefit from either strategy may enable personalized testing.

Methods and results

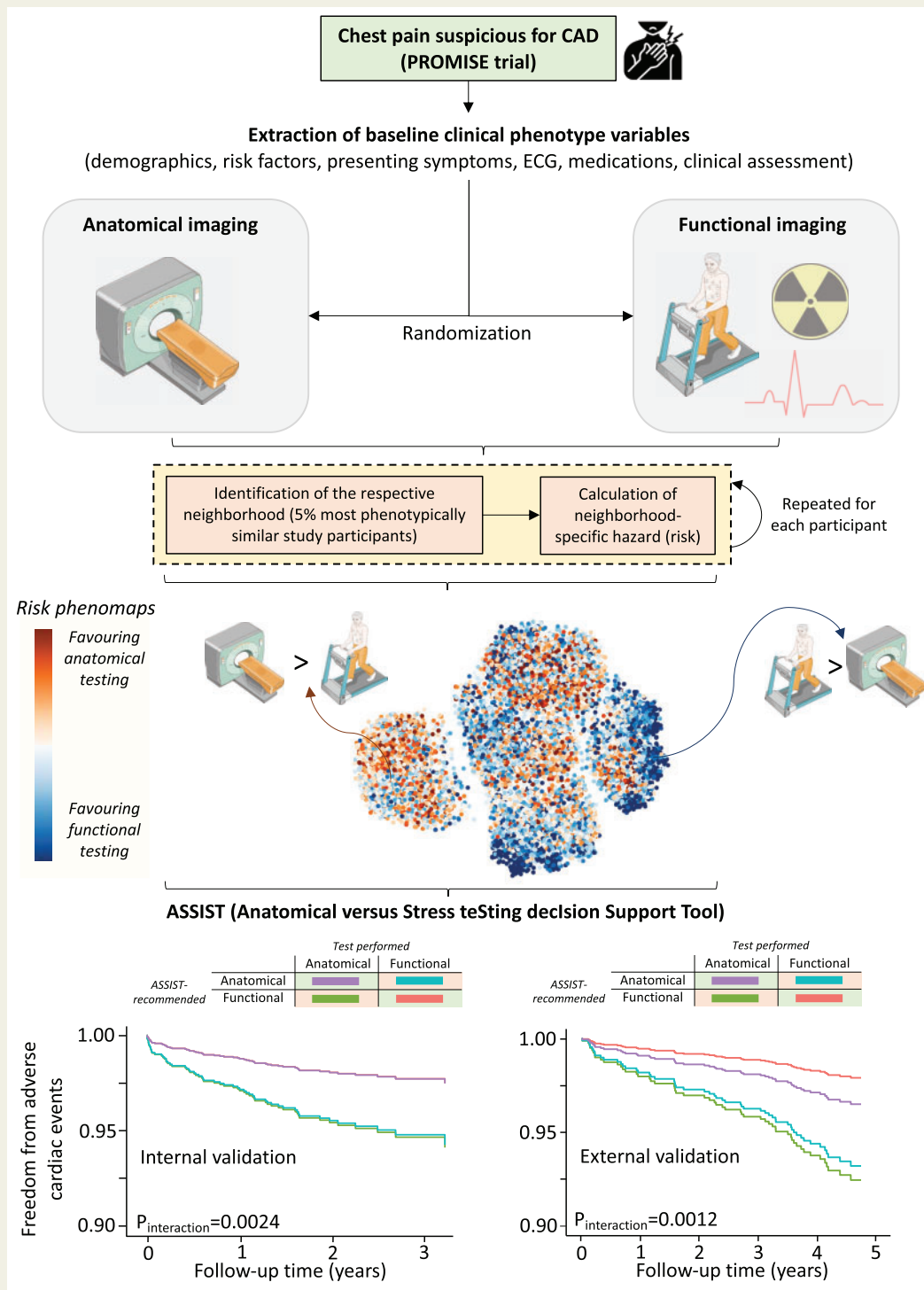
Using participant-level data from 9572 patients undergoing anatomical ($n = 4734$) vs. functional ($n = 4838$) testing in the PROMISE (PROspective Multicenter Imaging Study for Evaluation of Chest Pain) trial, we created a topological representation of the study population based on 57 pre-randomization variables. Within each patient's 5% topological neighbourhood, Cox regression models provided individual patient-centred hazard ratios for major adverse cardiovascular events and revealed marked heterogeneity across the phenomap [median 1.11 (10th to 90th percentile: 0.52–2.61)], suggestive of distinct phenotypic neighbourhoods favouring anatomical or functional testing. Based on this risk phenomap, we employed an extreme gradient boosting algorithm in 80% of the PROMISE population to predict the personalized benefit of anatomical vs. functional testing using 12 model-derived, routinely collected variables and created a decision support tool named ASSIST (Anatomical vs. Stress teSting decision Support Tool). In both the remaining 20% of PROMISE and an external validation set consisting of patients from SCOT-HEART (Scottish COmputed Tomography of the HEART Trial) undergoing anatomical-first vs. functional-first assessment, the testing strategy recommended by ASSIST was associated with a significantly lower incidence of each study's primary endpoint ($P = 0.0024$ and $P = 0.0321$ for interaction, respectively), as well as a harmonized endpoint of all-cause mortality or non-fatal myocardial infarction ($P = 0.0309$ and $P < 0.0001$ for interaction, respectively).

Conclusion

We propose a novel phenomapping-derived decision support tool to standardize the selection of anatomical vs. functional testing in the evaluation of stable chest pain, validated in two large and geographically diverse clinical trial populations.

* Corresponding author. Tel: 203-764-5885, Email: rohan.khera@yale.edu

Graphical Abstract



A novel phenomapping approach, trained and validated in patients from two large randomized clinical trials, evaluated the clinical value of coronary computed tomography vs. functional testing to individualize the selection of the appropriate diagnostic test for stable chest pain. CAD, coronary artery disease; ECG, electrocardiogram.

Keywords

Chest pain • Phenomapping • Machine learning • Computed tomography • Stress testing

Introduction

Nearly 200 million people globally suffer from coronary artery disease (CAD),¹ half of whom initially present with chest pain.² The optimal non-invasive diagnostic strategy for chest pain in patients with suspected stable CAD is clinically important to define, yet remains uncertain.³ PROMISE (PROspective Multicenter Imaging Study for Evaluation of Chest Pain)⁴ and SCOT-HEART (Scottish COmputed Tomography of the HEART Trial)^{5,6} have recently demonstrated that anatomical imaging has comparable outcomes to stress testing and may improve long-term outcomes when used in addition to standard of care including stress testing. This allowed computed tomography coronary angiography (CTCA) to gain traction as an alternative to functional testing.^{7,8} However, the choice between these two strategies remains arbitrary, despite over 14 000 randomized individuals across these large, well-conducted trials. This clinical equipoise is evident in the recent European Society of Cardiology (ESC) guidelines that assign a Class I recommendation to both CTCA and non-invasive functional testing as appropriate initial tests to diagnose CAD in symptomatic patients.⁹

The PROMISE trial remains the largest randomized controlled trial to have compared CTCA with functional testing in low-risk symptomatic patients with stable chest pain,⁴ and included 10 003 individuals followed for a median 25 months.⁴ However, subsequent analyses have revealed evidence of heterogeneity across broad subgroups, with women compared with men, and patients with diabetes compared with those without diabetes experiencing fewer adverse cardiovascular events with anatomical testing than with functional testing.^{10–12}

Nevertheless, broad subgroup assessments do not account for large variation in demographic and clinical features within such subgroups. Therefore, there are no tools that support individualization of the expected benefit of anatomical and functional testing based on each patient's unique phenotype, which is essential for shared decision-making.

In this study, we developed a method that evaluates the phenotypic diversity of patients presenting with stable chest pain as well as their optimal non-invasive testing strategy based on each patient's unique set of pre-randomization characteristics, and subsequent outcomes, using individual patient data from two major clinical trials investigating the clinical value of anatomical testing in the evaluation of chest pain (*Graphical abstract*).

Methods

Data source

We obtained participant-level data of the PROMISE trial through the National Heart, Lung and Blood Institute. Details of the PROMISE trial have been previously published.⁴ Briefly, PROMISE (ClinicalTrials.gov identifier: NCT01174550) recruited 10 003 patients from multiple centres in the USA and Canada who were randomized to either anatomical (CTCA) or functional testing (including exercise electrocardiography, nuclear stress testing, or stress echocardiography).⁴ The Yale Institutional Review Board approved our study and waived the requirement for informed consent for our *post hoc* analysis of de-identified data. SCOT-HEART (ClinicalTrials.gov identifier: NCT01149590) enrolled 4146 patients from 122 cardiology chest pain clinics across Scotland and

randomized to CTCA in addition to standard care compared with standard care alone for the evaluation of stable chest pain, as previously described.^{5,6} The dataset was made available through a collaboration with the original study investigators. We confirm that the present study complied with the Declaration of Helsinki.

Study population and covariates

In PROMISE, we identified all individuals who underwent initial assessment with anatomical or functional testing, consistent with their original randomized assignment. This represented 9572 of the 10 003 original participants. We included patient characteristics available at trial enrolment, including demographics (age, sex, race, ethnicity), anthropometrics [body mass index (BMI)], cardiovascular risk factors (systolic and diastolic blood pressure, hypertension, diabetes mellitus, smoking status, family history), laboratory measurements (haemoglobin, creatinine, lipid panel), medications, presenting symptoms (i.e. chest pain, shortness of breath), chest pain characteristics (typical, atypical, non-cardiac), electrocardiographic parameters (i.e. rhythm, Q waves, findings interfering with stress test interpretation), and clinical risk scores (pooled cohort equation-derived 10-year atherosclerotic cardiovascular disease risk¹³ and modified Diamond-Forrester risk for obstructive CAD¹⁴). We excluded variables from model development if they were missing in over half of the participants (included/excluded variables listed in [Supplementary material online, Table S1](#)) or if they were recorded after study initiation.¹⁵ We imputed missing data for the included variables using chained random forests with predictive mean matching.¹⁶ Following imputation, we transformed continuous variables into standardized scores (z-scores) by subtracting their mean and dividing by their respective standard deviation ([Supplementary material online, Methods](#)).

In SCOT-HEART, we identified a subpopulation that underwent diagnostic evaluation of chest pain similar to PROMISE. This subpopulation included individuals in the CTCA arm who underwent anatomical testing without antecedent stress electrocardiogram (ECG) in clinic (anatomical-first arm). In the control arm that did not undergo CTCA, we identified all individuals with an initial stress test, across all modalities of stress ECG, radionuclide perfusion, stress echocardiography, or stress magnetic resonance imaging (functional-first arm).

Study outcomes

To ensure consistency with the original trials, we used each study's pre-specified primary endpoint. In PROMISE, our primary study population, we trained our models using a composite of death, myocardial infarction (MI), unstable angina hospitalization, or major procedural complication (major adverse cardiovascular events [MACE]).⁴ In SCOT-HEART, the primary endpoint was a composite of death due to coronary heart disease and non-fatal MI.⁶

Further, to harmonize the two studies for validation of our algorithm, we also identified a secondary composite endpoint of all-cause mortality and non-fatal MI that was captured in both studies.

Defining phenotypic neighbourhoods

In PROMISE, we computed a dissimilarity index that classified individuals based on 57 pre-randomization characteristics according to the Gower distance, a metric of dissimilarity between two patients based on mixed numeric and non-numeric data.¹⁷ For each patient in PROMISE, we identified a topological neighbourhood of the 5% most phenotypically similar participants based on Gower's distance. In sensitivity analyses, we iteratively evaluated random neighbourhood sizes between 2.5% and 10%, assessing the correlation of effect estimates in these iterations with those derived from the 5% neighbourhood size (see [Supplementary material online, Methods](#)).

Individualized risk phenomapping

Within each patient-centred neighbourhood, we assessed the association of undergoing anatomical vs. functional testing with MACE in age- and sex-adjusted Cox regression models, thus providing individualized risk estimates based on each patient's unique neighbourhood. The natural logarithmic transformations of the hazard ratio (HR) from the Cox models comparing anatomical and functional testing for each patient's topological neighbourhood represented their individualized effect estimate. In our approach, negative log-HRs favour anatomical testing, whereas positive values favour functional testing.

Furthermore, since an unbiased personalized effect estimate is contingent upon the similarity of individuals in their topological neighbourhoods, we created a measure of neighbourhood homogeneity. This represented the square of 1 minus the average pairwise distance between the index patient and each one of their neighbours, with higher values reflecting a neighbourhood of phenotypically more similar patients.

To visualize the phenotypic variation in the PROMISE population and neighbourhoods we used uniform manifold approximation and projection (UMAP),¹⁸ which constructs a two-dimensional representation of the high-dimensional feature space. We employed colour maps to visualize the topological distribution of the patient baseline demographics and neighbourhood estimates in the phenomap.

We demonstrated the ability of our approach to detect heterogeneity in treatment effects using examples of individuals sharing a key set of features (age, sex, traditional risk factors) but differing on other baseline characteristics.

Extreme gradient boosting algorithm to predict the benefit of anatomical testing

To translate the heterogeneity in treatment effect across the PROMISE phenomap to a clinical population, we constructed an extreme gradient boosting algorithm to predict the personalized risk of MACE with anatomical vs. functional testing (natural logarithm of the neighbourhood HR) using variables, which were available in $\geq 50\%$ of participants in both PROMISE and SCOT-HEART, spanning demographics, comorbidities, laboratory testing, vitals, and medications with implications for anatomical or functional testing. This process yielded 21 variables including key demographics (age, sex), risk factors (smoking, family history of CAD, hypertension, diabetes mellitus, total cholesterol, high-density lipoprotein, statin use), anthropometrics (BMI, systolic and diastolic blood pressure), cerebrovascular and peripheral vascular disease, ECG findings (rhythm, Q waves, findings interfering with stress test interpretation, as defined in PROMISE),⁴ and use of antiplatelets and beta-blockers.

We randomly divided the PROMISE population into training (80%, $n = 7660$) and internal validation (20%, $n = 1912$) sets. Data pre-processing, parameter set-up, and hyperparameter tuning are described in the [Supplementary material online, Methods](#). Briefly, we trained the extreme gradient boosting algorithm to identify patient characteristics that were strongly associated with improved outcomes (patient-centred log-hazards) for anatomical or functional testing. We used root mean squared error to evaluate our model performance, identified the optimal hyperparameters using a grid search, and implemented 10-fold cross-validation. We evaluated feature importance using SHAP (SHapley Additive exPlanations) values,¹⁹ which identify a predictor contribution, either positively or negatively, to the prediction.

To improve the model practical application, we selected features that were strongly associated with improved outcomes with either anatomical or functional testing based on a feature importance of 0.03 or higher, resulting in 12 features. We retrained our model using these limited set of features, using 10-fold cross-validation in the 80% of PROMISE, followed by further validation in the remaining (unseen) 20% of PROMISE.

This machine learning-derived parsimonious model trained on 12 features represented ASSIST (Anatomical vs. Stress teSting decision Support Tool). Negative ASSIST values (<0) predicted improved outcomes with anatomical-first assessment, whereas positive ASSIST values (>0) favoured functional-first assessment.

External validation and performance of ASSIST

We validated the decision support tool, ASSIST, externally in a selected subset of SCOT-HEART that underwent anatomical-first vs. functional-first assessment. As the SCOT-HEART subset represented a selected population with the potential for intervention and covariate imbalance, in a sensitivity analysis, we created propensity score-matched subgroups of patients undergoing anatomical or functional testing ([Supplementary material online, Methods](#)).

Statistical analyses

We compared the two treatment groups using Student's *t*-test for continuous variables and chi-square test for categorical variables and used Pearson's correlation to assess continuous variables. We performed survival analyses using Cox proportional-hazards regression. While neighbourhoods were matched on pre-randomization covariates, we explicitly adjusted Cox models for age and sex. We assessed the association of the ASSIST recommended testing modality and outcomes through its group-wise interactions with the two treatment groups in Cox models. Statistical tests were two-sided with a level of significance of 0.05. Analyses were performed using R (version 4.0.2) and Python (version 3.8.5). Reporting of the study design and findings stands consistent with the STROBE (Strengthening the reporting of observational studies in epidemiology) guidelines (see checklist in [Supplementary material online, Data](#)).²⁰

Role of the funding source

Funding sources had no involvement in the study design, collection, analysis, interpretation of the data, or the decision to submit the paper for publication.

Results

Study population

From PROMISE, we included 9572 patients [age 60.3 ± 8.3 years, $n = 5013$ (52.4%) women] with stable chest pain. Of these, 4734 (49.5%) underwent CTCA and the remaining 4838 (50.5%) functional testing (*Figure 1A*). Baseline characteristics were balanced between the two study arms ([Supplementary material online, Table S2](#)). Over a mean follow-up period of 2.1 ± 0.9 years, there were 294 MACE (primary study outcome), with no significant difference in the primary outcome in the two arms [adjusted HR 1.03 (95% confidence interval (CI): 0.82–1.29), $P = 0.8159$ for anatomical vs. functional testing].

In the external validation dataset from SCOT-HEART, 2135 patients [age 57.1 ± 9.8 years, $n = 949$ (44.4%) women] underwent either anatomical testing alone ($n = 301$) or functional testing alone ($n = 1834$) (*Figure 1B*, baseline characteristics in [Supplementary material online, Table S3](#)). Over an average follow-up of 4.8 ± 1.1 years, a total of 74 primary outcome events of coronary heart disease death or non-fatal MI events were recorded [adjusted HR 0.63 (95% CI 0.29–1.37), $P = 0.2411$], an effect size consistent with the beneficial

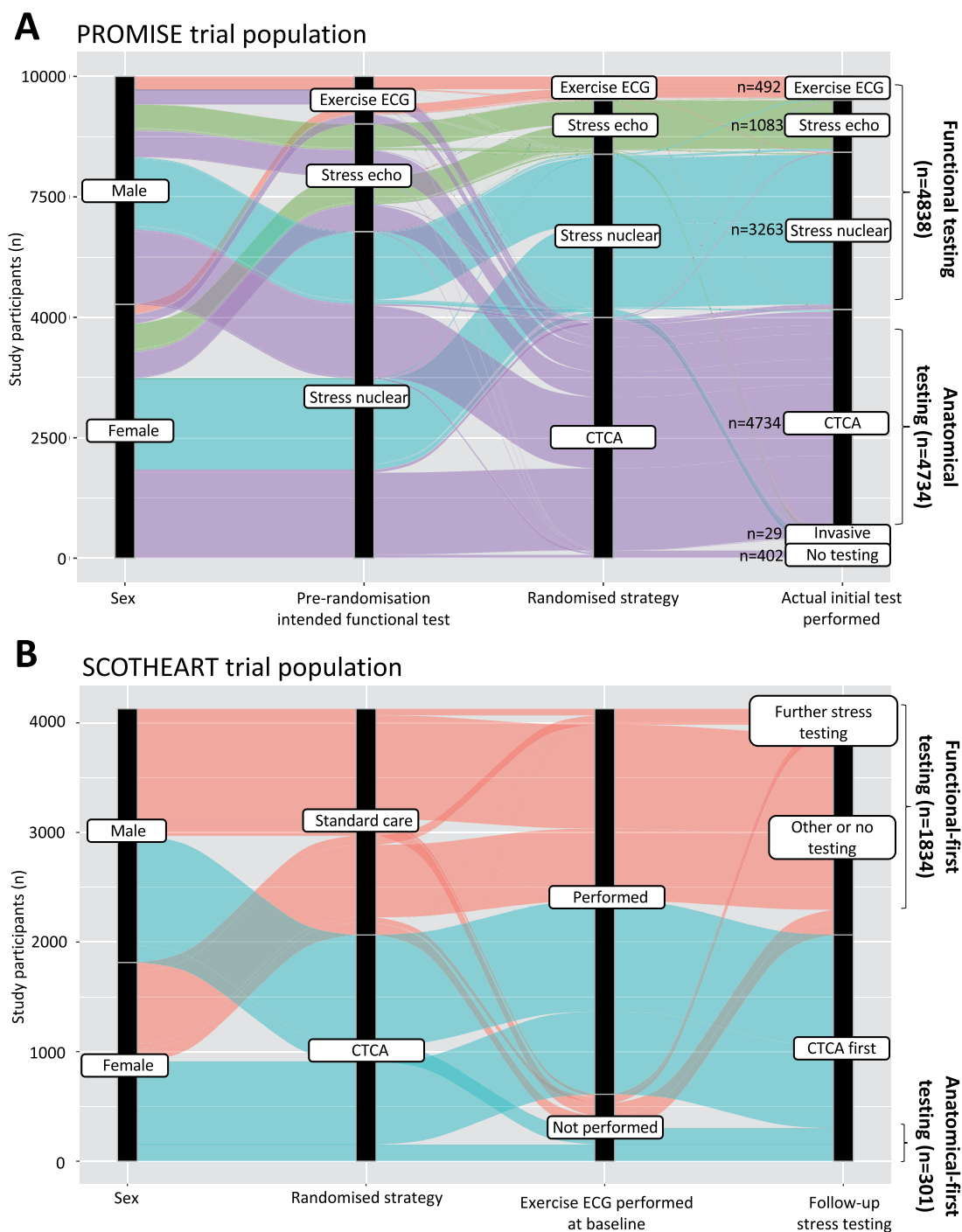


Figure 1 Alluvial diagram of diagnostic testing in PROMISE and SCOT-HEART. (A) Among 10 003 participants randomized to anatomical vs. functional testing in the PROMISE trial, a total of 4834 vs. 4734 individuals underwent an anatomical vs. functional test as their initial investigation (and were included in this study), with 402 patients receiving no testing and the remaining 29 undergoing invasive coronary angiography as the initial diagnostic test. (B) Among 4146 chest pain patients in SCOT-HEART who were randomized to standard care vs. standard care plus computed tomography coronary angiography, a total of 301 patients underwent computed tomography coronary angiography imaging without baseline stress electrocardiography vs. 1834 patients who underwent stress/functional testing as their initial investigation without the addition computed tomography coronary angiography. CTCA, computed tomography coronary angiography; ECG, electrocardiography; PROMISE, PROspective Multicenter Imaging Study for Evaluation of Chest Pain; SCOT-HEART, Scottish COmputed Tomography of the HEART Trial.

role of CTCA seen in the original trial ([Supplementary material online, Figure S1](#)).

Phenomapping the stable chest pain in PROMISE

We first created a phenomap of our study population using a pairwise dissimilarity metric derived from 57 pre-randomization phenotypic characteristics and visualized it as a two-dimensional manifold representation. Based on visual assessment, the two treatment arms were distributed uniformly throughout the phenotypic space, consistent with their random allocation across the population ([Figure 2A](#)) with varying baseline clinical factors and risk of CAD ([Figure 2B](#)).

Distribution of neighbourhood-based individualized risk estimates

Patient-specific neighbourhoods for each of the 9572 included PROMISE participants, included 5% of the population in their topological vicinity, with a wide distribution of neighbourhood-specific risk effect estimates ([Supplementary material online, Figure S2](#)). The median neighbourhood-specific HR for MACE was 1.11 with 10th, 25th, 75th, and 90th percentiles of 0.52, 0.76, 1.67, and 2.61, respectively. A projection of each person's individual effect estimate on the phenomap suggested distinct topological neighbourhoods favouring anatomical or functional testing ([Figure 2C and D](#)). There was also variation in both the direction of the effect and the effect size for different endpoints across the topological space of the study population.

In sensitivity analyses for variable neighbourhood sizes (2.5%, 5%, 7.5%, 10%, 15% of the study population), an increasing neighbourhood size was associated with a narrower distribution of individual risk estimates around the average treatment effect across the cohort (HR 1.03), representing loss of risk heterogeneity observed at larger neighbourhood sizes. The larger neighbourhood, however, also compared dissimilar individuals with decreasing neighbourhood homogeneity based on increasing mean distances ([Supplementary material online, Figure S3](#)). Random iterations for various neighbourhood sizes between 2.5% and 10% showed that the average effect size was strongly correlated with that derived from 5% neighbourhoods [$r = 0.72$ (95% CI 0.71–0.73)] ([Supplementary material online, Figure S4](#)).

Using risk phenomap for individualized risk prediction

To demonstrate an example of individualized risk estimation using the phenomap, we identified a subset of three phenotypically similar PROMISE participants, each of them a 59-year-old woman, with a history of diabetes and hypertension but not smoking, presenting with atypical chest pain and a modified pre-test Diamond-Forrester score of 20%. Despite the above similarities, phenomapping using all 57 included variables ([Supplementary material online, Table S4](#)) revealed that these patients were located in distinct topological neighbourhoods ([Figure 3A](#)). Each patient's neighbourhood-specific assessments identified differential risk/benefit associated with anatomical vs. functional testing, ranging from improved outcomes with functional testing ([Figure 3B](#)) to similar outcomes with either strategy ([Figure 3C](#)), or improved outcomes with anatomical imaging ([Figure 3D](#)). Of note,

each patient neighbourhood had phenotypically similar patients in the two study arms ([Supplementary material online, Table S5](#)).

The Anatomical vs. Stress testing decision Support Tool (ASSIST)

In the 80% training set from PROMISE ($n = 7660$), an extreme gradient boosting algorithm identified hypertension, diabetes mellitus, use of beta-blockers, female sex, statin use, smoking history, antiplatelet use, BMI, age, and cholesterol levels as the predictors with highest feature importance for relative hazard of MACE with anatomic or functional testing ([Figure 4A](#)). Feature importance analysis suggested that female sex, hypertension, diabetes mellitus, use of beta-blockers, and active or former smoking were each associated with improved outcomes with anatomical testing ([Figure 4B](#)), whereas absence of these risk factors as well as lower BMI and statin use favoured functional testing. Our clinical decision support tool, ASSIST, represents the extreme gradient model developed using these 12 most important features. Hold-out validation performance of the parsimonious 12-feature tool was comparable with that of a model relying on all 21 inputs (RMSE of 0.59 vs. 0.57, respectively), while logistically easier to deploy.

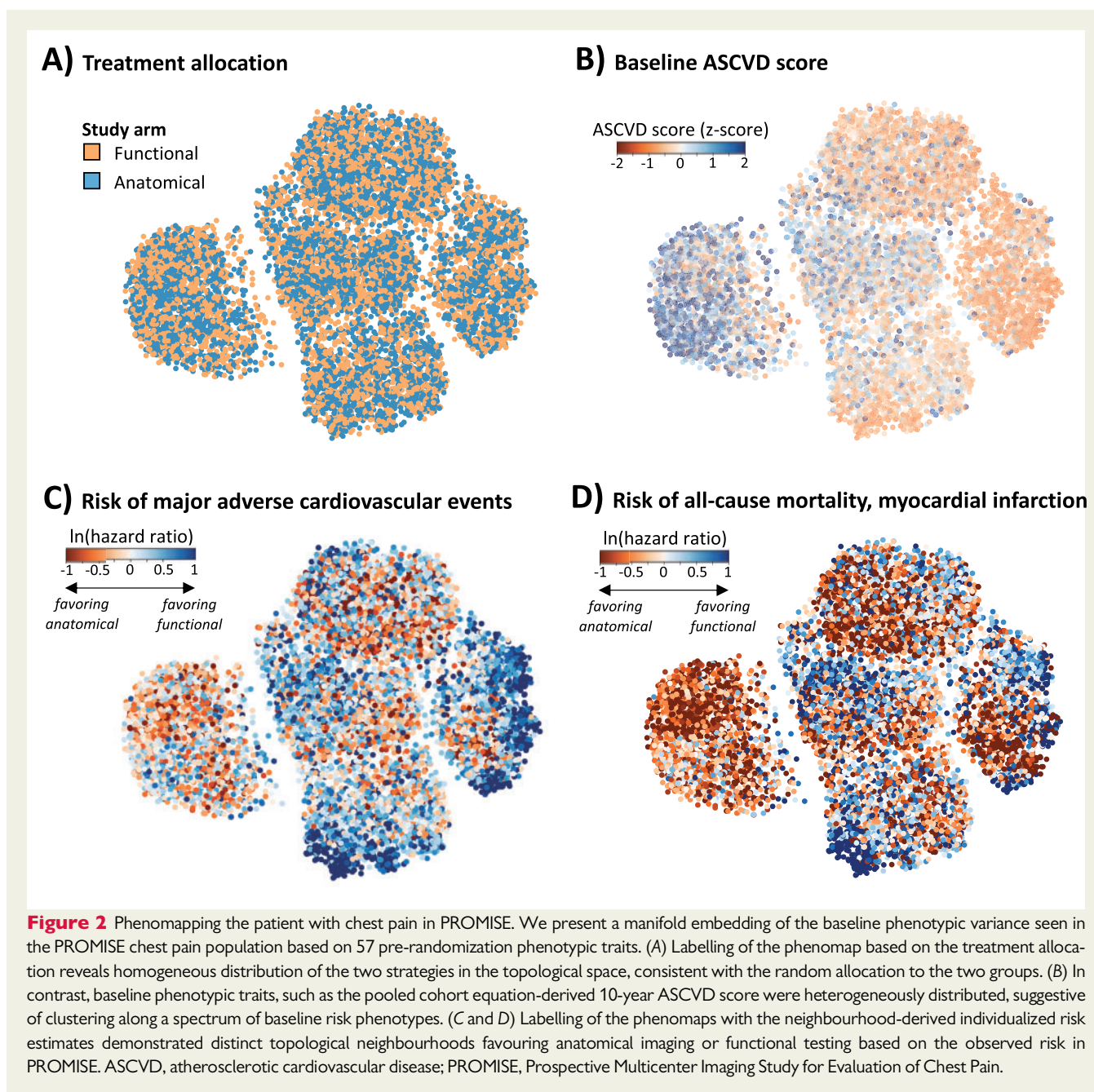
Of note, in both the cross-validated training and testing sets of PROMISE, there was no association between the ASSIST risk prediction and the allocation to either anatomical or functional testing, consistent with the random allocation to the two arms ([Figure 4C and D](#)).

Internal and external validation of ASSIST

In the remaining 20% PROMISE participants ($n = 1912$, *internal validation*; [Figure 5](#)), as well as in the selected unmatched and propensity score-matched population of SCOT-HEART (*external validation*; [Figure 6](#)), the ASSIST performed well in identifying the favoured diagnostic strategy. In both the internal and external validation sets, agreement between the ASSIST recommendation (score >0: favouring functional, score <0: favouring anatomical) and the actual test performed was associated with a significantly lower incidence of each primary composite endpoint ([Figures 5A–C and 6A–C](#), respectively) as well as the harmonized endpoint of all-cause mortality and non-fatal MI ([Figures 5D–F and 6D–F](#), respectively), with consistent significant interaction between the ASSIST-recommended test and performed strategy ([Figures 5 and 6](#)). Findings demonstrated a dose–response relationship in a pooled analysis of the two cohorts ([Supplementary material online, Figure S5](#)) and were also robust to a sensitivity analysis for all-cause mortality ([Supplementary material online, Figure S6](#)). Of note, a *post hoc* analysis of individual risk factors in the external validation set did not identify patients more likely to have favourable outcomes with anatomical vs. functional testing ($P_{\text{interaction}} = 0.7925$ for sex, 0.3450 for hypertension, and 0.8474 for diabetes mellitus).

Discussion

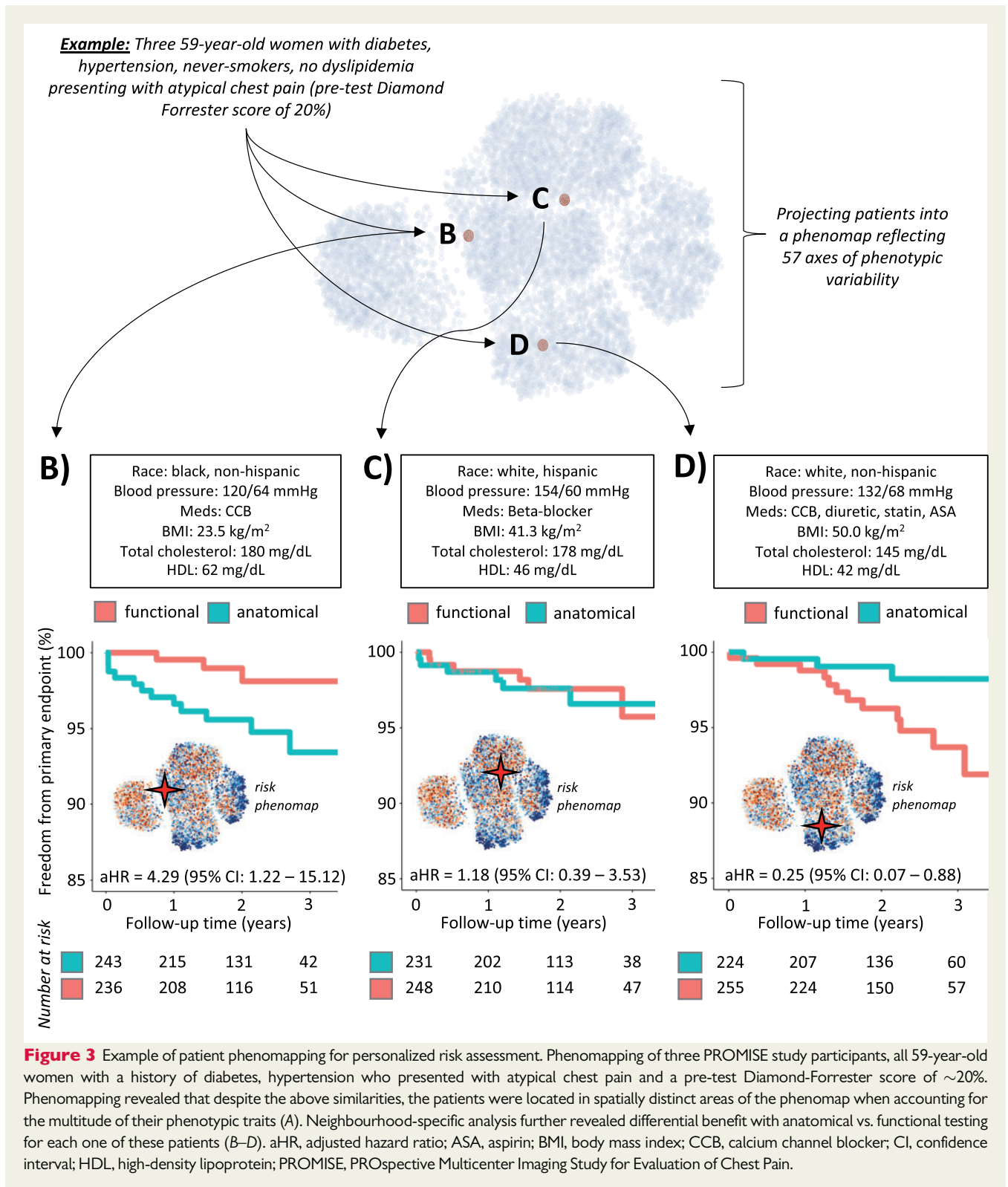
In the largest clinical trial to have evaluated the role of CTCA in the investigation of stable chest pain, we developed and validated a machine learning-based decision support tool to guide the selection between anatomical and functional evaluation. We defined a novel



strategy that constructs a high-dimensional phenotypic representation of trial participants, permitting a series of local experiments within the trial uncovering heterogeneous treatment effects, identifying individuals who may derive benefit from one strategy over another. Our approach synthesizes the complex relationship between a large number of pre-randomization characteristics in creating and visualizing a comprehensive phenomap of patients, with an individualized assessment of the risk of adverse cardiovascular events with anatomical or functional testing for assessing chest pain. Our new machine learning-derived tool (ASSIST) based on 12 widely available clinical parameters derived from risk phenomaps reliably and consistently identified patients who were more

likely to have improved outcomes when assigned to an anatomical or functional diagnostic strategy.

To date, there has been no consensus on the strategy to choose between anatomical and functional testing in chest pain evaluation,²¹ and different clinical practice guidelines provide varying levels and strengths of recommendation on the use of CTCA vs. functional testing.^{8,9,22,23} Despite PROMISE and SCOT-HEART, identifying a population that may benefit from CTCA or functional testing has been mostly supported through *post hoc* analyses in large population subgroups, specifically women,¹⁰ and patients with diabetes,¹¹ and considerations about CTCA test characteristics, including high sensitivity,²⁴ but limited specificity in detecting haemodynamically



significant lesions.²⁵ Therefore, the default strategy may be to use CTCA in individuals at presumably low-to-intermediate risk of CAD.⁹ Unfortunately, this approach does not benefit from the knowledge gained from the large clinical trials and the extensive phenotypic variability among trial participants. Our approach overcomes these

limitations through a specific focus on a large feature set and their complex relationship to each other, therefore deriving a personalized estimate, as opposed to an average treatment effect across large heterogeneous groups. In addition, instead of focusing on the absolute risk of obstructive disease or myocardial ischaemia, our study

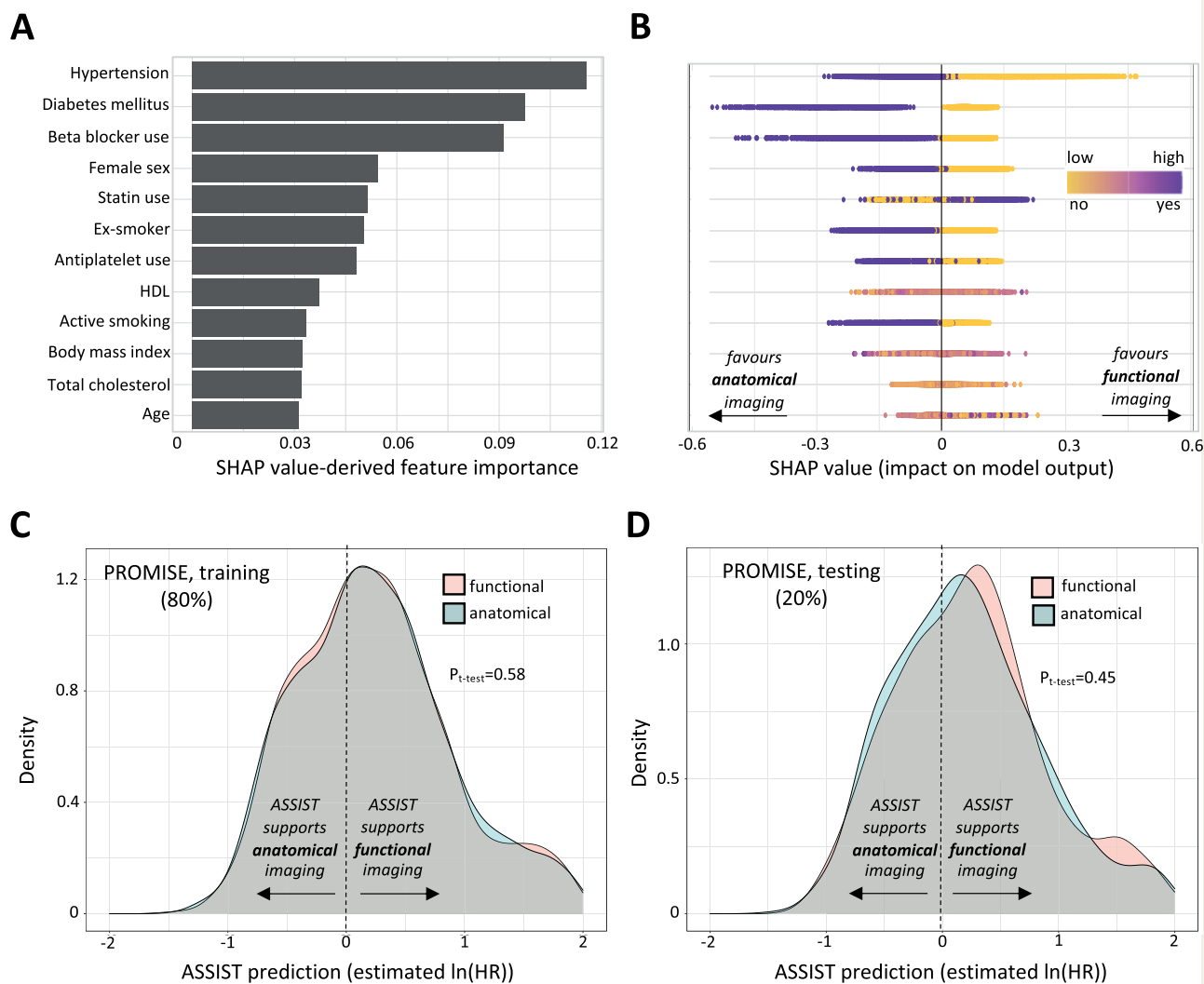


Figure 4 Developing a decision support tool to predict individualized benefit from anatomical vs. functional testing in chest pain investigation. (A) In a randomly selected sample of the PROMISE population, we trained an extreme gradient boosting tree to predict the phenomap-derived individualized risk with anatomical vs. functional testing. We identified the most important input features based on the SHAP values and selected the top 12 predictors (all with feature importance of 0.03 or higher) to create an easy-to-use clinical support tool, named ASSIST. (B) To offer some insight into each variable contribution, we used a SHAP summary plot, in which the y-axis represents the variables in descending order of importance and the x-axis indicates the change in prediction. The gradient colour denotes the original value for that variable (for instance for Booleans such as hypertension or diabetes it only takes two colours, whereas for continuous variables it contains the whole spectrum), with each point representing an individual from the original training set. Negative SHAP values (x-axis) indicate improved outcomes with anatomical imaging (as seen among individuals with hypertension and diabetes) whereas positive values indicate improved outcomes with functional testing. (C and D) Notably, ASSIST predictions were independent of the random assignment to the anatomical or functional testing group in both the training and testing sets of PROMISE. ASSIST, Anatomical vs. Stress teSting decision Support Tool; PROMISE, PROspective Multicenter Imaging Study for Evaluation of Chest Pain; SHAP, SHapley Additive exPlanations.

explores the factors associated with the relative benefit obtained from anatomical vs. functional testing.

Our study uses a novel approach to achieve these goals. Our approach leverages the detailed phenotypic characterization of clinical trial populations at enrolment and the unbiased treatment allocation to infer a personalized treatment effect. Therefore, it provides a quantitative evaluation of the heterogeneity of outcomes, and an assessment whether the average treatment effect

observed in a clinical trial setting applies to a given trial participant. We also created a visual representation of differences across individuals enrolled in a clinical trial, allowing interpretability of different patients and the observed effects. Our approach builds upon prior studies that have employed clustering to demonstrate clinical trial participants have discordant effects.^{26–29} However, they are limited in clinical application as they ultimately represent broad subgroups of patients that differ from each other on many

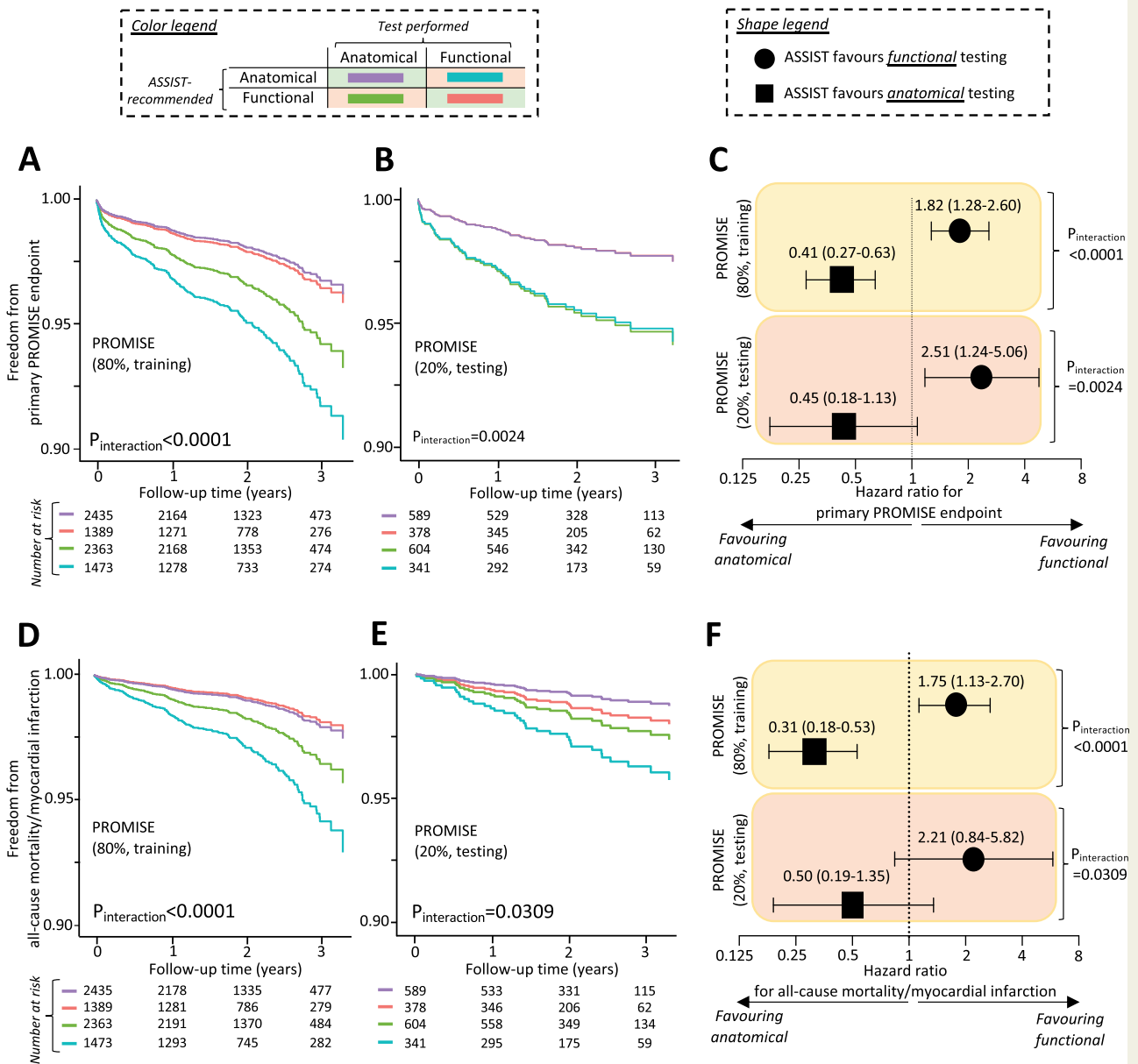


Figure 5 Internal validation and performance of ASSIST in PROMISE. Application of the ASSIST tool in both the training and testing (internal validation) set of PROMISE demonstrated that concordance (vs. disagreement) between the ASSIST-proposed best initial diagnostic strategy and a patient random allocation to functional or anatomical testing was associated with an approximate two-fold reduction in the risk of the study primary composite endpoint (A–C), as well as a composite endpoint of all-cause mortality and non-fatal myocardial infarction (D and E). ASSIST, Anatomical vs. Stress teSting decision Support Tool; PROMISE, PROspective Multicenter Imaging Study for Evaluation of Chest Pain.

characteristics, thereby limiting a personalized treatment selection. In our approach, each individual represents the centre of their own cluster and, therefore, is compared with similar individuals in inferring a treatment effect.

In addition, our machine learning-based decision support tool, ASSIST, allows such personalization of the diagnostic strategy for chest pain using only 12 key variables. The tool (which is available online,³⁰ also see [Supplementary material online, Figure S7](#)) consistently

demonstrated a lower rate of all-cause mortality and adverse cardiovascular outcomes where the diagnostic strategy was aligned with ASSIST recommendation. Moreover, these findings were replicated in the SCOT-HEART study, which included a geographically and phenotypically distinct population. Notably, previously suggested broad demographic and clinical groups that may benefit from anatomical evaluation^{10,11} were not generalizable to the external population.

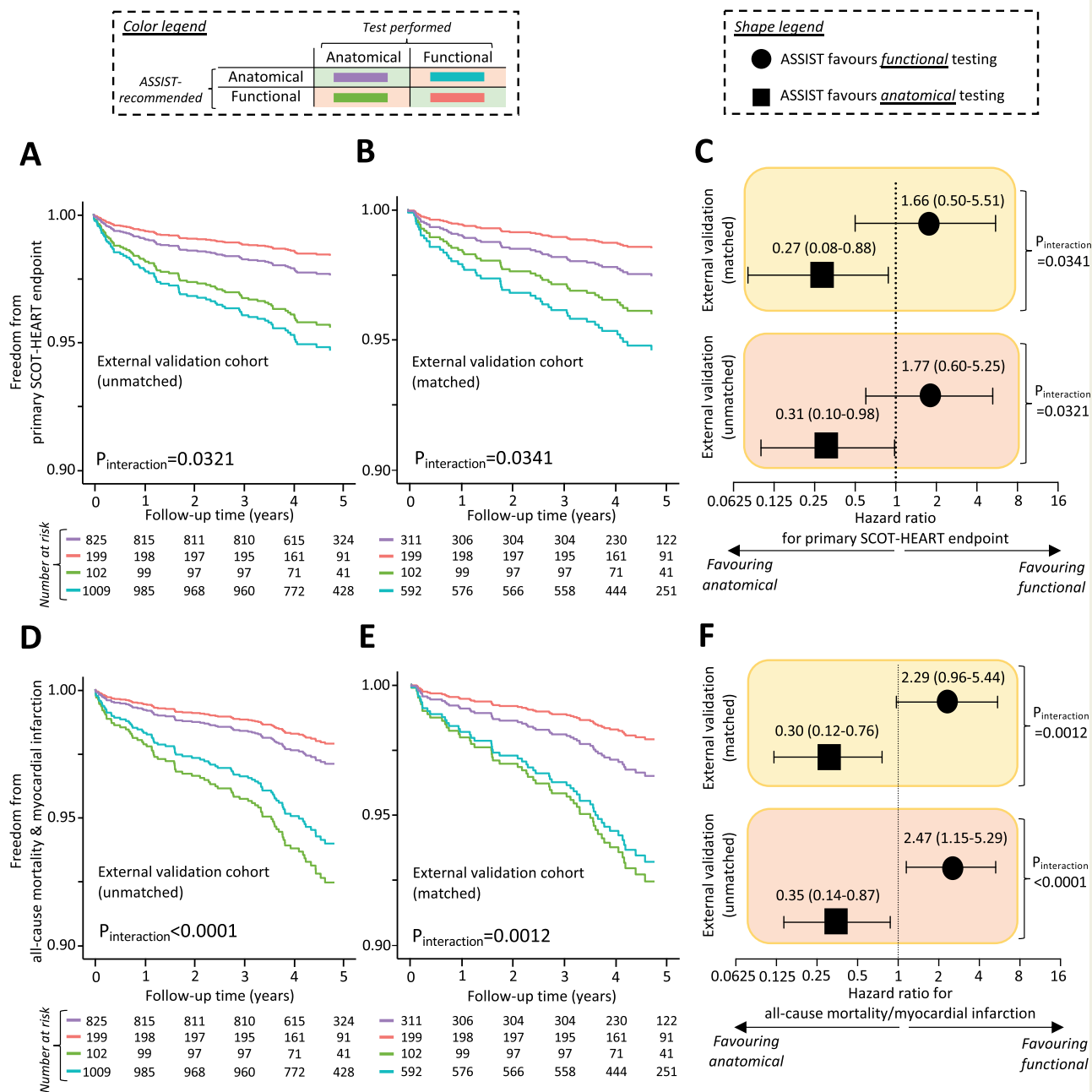


Figure 6 External validation and performance of ASSIST. Application of the ASSIST tool in both the unmatched and matched subpopulation of SCOT-HEART used for external validation purposes confirmed that concordance (vs. disagreement) between the ASSIST-proposed best initial diagnostic strategy and a patient allocation to anatomical-first imaging vs. standard of care (which included functional-first testing) was associated with a reduction in the risk of the study primary composite endpoint (A–C), as well as the harmonized composite endpoint of all-cause mortality and non-fatal myocardial infarction (D and E). ASSIST, Anatomical vs. Stress teSting decision Support Tool; SCOT-HEART, Scottish COmputed Tomography of the HEART Trial.

Our study should be interpreted in light of the following limitations. First, our approach does not offer a mechanistic explanation for the differential effect of testing strategy on outcomes. Both trials did not require a fixed treatment cascade that follows each strategy, and differences in clinical endpoints likely reflect medical and

procedural interventions that follow these tests. A validation in two distinct cohorts suggests consistency in the observation and generalizability of our findings; however, treatment strategies deviating from PROMISE and SCOT-HEART may not find a similar effect. Second, our study is a *post hoc* analysis, and therefore, our findings are

descriptive and only represent a way to interpret trial data. While randomization was not fully preserved, the treatment allocation itself was not driven by clinical indications and diagnostic strategies were assigned as part of the trial protocol. This overcomes the limitation of other observational studies based on real-world data, which are often subject to confounding by indication. However, our findings do highlight heterogeneity in risks and benefits of anatomical and functional testing, and a significant interaction between the test received and a patient baseline profile suggests that ASSIST accurately identified individuals for whom anatomical or functional testing may have differential outcome implications. Therefore, ASSIST may aid patient discussions, particularly at centres where both tests are similarly accessible. To this end, future validation in unselected cohorts will be crucial. Third, SCOT-HEART had a different design than PROMISE, with the majority (~85%) of patients in the computed tomography coronary angiography arm undergoing baseline exercise ECG (rather than nuclear imaging), whereas as the standard of care control arm did not explicitly require functional testing.⁵ We harmonized the diagnostic testing in SCOT-HEART to PROMISE to allow its use as a validation study, and explicitly conducted a matched analysis of SCOT-HEART participants with equal propensity of undergoing anatomical or functional testing as their first test. However, this process resulted in loss of randomization and the conclusions of the study would not be generalizable to the strategy assessed in the overall SCOT-HEART population, which demonstrated the benefit of anatomical imaging when used in addition to rather than as an alternative to functional testing.^{5,6} Finally, while missing data were imputed only if missing in a minority of the included patients, imputation may limit the accuracy of some of our study estimates.

Conclusion

We have developed an approach that defines an evidence-based strategy to pursue anatomical or functional evaluation of patients with suspected CAD. The approach uses a series of local experiments in a multidimensional phenomap of trial participants to infer a personalized strategy of the diagnostic evaluation approach most likely to achieve the best outcomes. Furthermore, a generalizable decision support tool derived from this phenomap, and validated in two geographically distinct large studies, enables a broader use of this information in shared decision-making in clinical practice.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Funding

The study was supported through a Yale School of Medicine grant to R.K. The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Acknowledgements

We thank Dr David E. Newby, the Principal Investigator of SCOT-HEART, for access to the trial data.

Data availability

The PROMISE clinical trial data are managed by the National Heart, Lung and Blood Institute's Biologic Specimen and Data Repository Information Coordinating Center and can be made available through a research application.

Conflict of interest: R.K. and E.K.O. report a planned patent relating to this work. E.K.O. has served as a consultant for Caristo Diagnostics, reports a patent GB2018/1818049.9 pending, a patent GB2016/1620494.3 pending, outside the scope of this work. M.A.S. reports grants from US National Institutes of Health, US Department of Veterans Affairs, US Food & Drug Administration, and Janssen Research & Development, personal fees from IQVIA and Private Health Management, outside the submitted work. J.d.L. reports serving on data monitoring committees for Novo Nordisk, Eli Lilly, Amgen, and Regeneron. C.A. reports grants from the British Heart Foundation, the National Institute of Health Research and Oxford Biomedical Research Centre, and Innovate UK, during the conduct of the study; personal fees and other from Caristo Diagnostics, outside the submitted work; a patent GB2015/052359 licensed to Caristo Diagnostics, a patent GB2016/1620494.3 licensed to Caristo Diagnostics, a patent GB2018/1818049.9 pending, a patent GR2018/0100490 pending, and a patent GR2018/0100510 pending. E.J.M. reports grants and consultancies from Eidos, Inc., Pfizer, Inc., and Alnylam, Inc., all outside the submitted work. The remaining authors have nothing to disclose.

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020;**396**:1204-1222.
2. Davies SW. Clinical presentation and diagnosis of coronary artery disease: stable angina. *Br Med Bull* 2001;**59**:17-27.
3. Poon M, Lesser JR, Biga C, Blankstein R, Kramer CM, Min JK, Noack PS, Farrow C, Hoffman U, Murillo J, Nieman K, Shaw LJ. Current evidence and recommendations for coronary CTA first in evaluation of stable coronary artery disease. *J Am Coll Cardiol* 2020;**76**:1358-1362.
4. Douglas PS, Hoffmann U, Patel MR, Mark DB, Al-Khalidi HR, Cavanaugh B, Cole J, Dolor RJ, Fordyce CB, Huang M, Khan MA, Kosinski AS, Krucoff MW, Malhotra V, Picard MH, Udelson JE, Velazquez EJ, Yow E, Cooper LS, Lee KL; PROMISE Investigators. Outcomes of anatomical versus functional testing for coronary artery disease. *N Engl J Med* 2015;**372**:1291-1300.
5. SCOT-HEART Investigators. CT coronary angiography in patients with suspected angina due to coronary heart disease (SCOT-HEART): an open-label, parallel-group, multicentre trial. *Lancet* 2015;**385**:2383-2391.
6. Newby DE, Adamson PD, Berry C, Boon NA, Dweck MR, Flather M, Forbes J, Hunter A, Lewis S, MacLean S, Mills NL, Norrie J, Roditi G, Shah ASV, Timmis AD, van Beek EJR, Williams MC; SCOT-HEART Investigators. Coronary CT angiography and 5-year risk of myocardial infarction. *N Engl J Med* 2018;**379**:924-933.
7. Foy AJ, Dhruva SS, Peterson B, Mandrola JM, Morgan DJ, Redberg RF. Coronary computed tomography angiography vs functional stress testing for patients with suspected coronary artery disease: a systematic review and meta-analysis. *JAMA Intern Med* 2017;**177**:1623-1631.
8. Moss AJ, Williams MC, Newby DE, Nicol ED. The updated NICE guidelines: cardiac CT as the first-line test for coronary artery disease. *Curr Cardiovasc Imaging Rep* 2017;**10**:15.
9. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, Prescott E, Storey RF, Deaton C, Cuisset T, Agewall S, Dickstein K, Edvardsen T, Escaned J, Gersh BJ, Svtil P, Gilard M, Hasdai D, Hatala R, Mahfoud F, Masip J, Muneretto C, Valgimigli M, Achenbach S, Bax JJ; ESC Scientific Document Group. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;**41**:407-477.
10. Pagidipati NJ, Hemal K, Coles A, Mark DB, Dolor RJ, Pelliikka PA, Hoffmann U, Litwin SE, Udelson J, Daubert MA, Shah SH, Martinez B, Lee KL, Douglas PS. Sex differences in functional and CT angiography testing in patients with suspected coronary artery disease. *J Am Coll Cardiol* 2016;**67**:2607-2616.

11. Sharma A, Coles A, Sekaran NK, Pagidipati NJ, Lu MT, Mark DB, Lee KL, Al-Khalidi HR, Hoffmann U, Douglas PS. Stress testing versus CT angiography in patients with diabetes and suspected coronary artery disease. *J Am Coll Cardiol* 2019;**73**:893–902.
12. Hernal K, Pagidipati NJ, Coles A, Dolor RJ, Mark DB, Pellicka PA, Hoffmann U, Litwin SE, Daubert MA, Shah SH, Ariani K, Bullock-Palmer RP, Martinez B, Lee KL, Douglas PS. Sex differences in demographics, risk factors, presentation, and noninvasive testing in stable outpatients with suspected coronary artery disease: insights from the PROMISE trial. *JACC Cardiovasc Imaging* 2016;**9**:337–346.
13. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB Sr, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith SC Jr, Sorlie P, Stone NJ, Wilson PWF. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014;**63**:2935–2959.
14. Genders TS, Steyerberg EW, Alkadhi H, Leschka S, Desbiolles L, Nieman K, Galema TW, Meijboom WB, Mollet NR, de Feyter PJ, Cademartiri F, Maffei E, Dewey M, Zimmermann E, Laule M, Pugliese F, Barbagallo R, Sinitsyn V, Bogaert J, Goetschalckx K, Schoepf UJ, Rowe GW, Schuijf JD, Bax JJ, de Graaf FR, Knuuti J, Kajander S, van Mieghem CA, Meijs MF, Cramer MJ, Gopalan D, Feuchtner G, Friedrich G, Krestin GP, Hunink MG; CAD Consortium. A clinical prediction rule for the diagnosis of coronary artery disease: validation, updating, and extension. *Eur Heart J* 2011;**32**:1316–1330.
15. Hummel M, Edelmann D, Kopp-Schneider A. Clustering of samples and variables with mixed-type data. *PLoS One* 2017;**12**:e0188274.
16. Wright MN, Ziegler A. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Soft* 2017;**77**:1–17.
17. Gower JC. A general coefficient of similarity and some of its properties. *Biometrics* 1971;**27**:857–871.
18. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426v3 [stat.ML]. 2018.
19. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;**2**:56–67.
20. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;**370**:1453–1457.
21. Fordyce CB, Newby DE, Douglas PS. Diagnostic strategies for the evaluation of chest pain: clinical implications from SCOT-HEART and PROMISE. *J Am Coll Cardiol* 2016;**67**:843–852.
22. Fihn SD, Blankenship JC, Alexander KP, Bittl JA, Byrne JG, Fletcher BJ, Fonarow GC, Lange RA, Levine GN, Maddox TM, Naidu SS, Ohman EM, Smith PK. 2014 ACC/AHA/AATS/PCNA/SCAI/STS focused update of the guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines, and the American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol* 2014;**64**:1929–1949.
23. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, Douglas PS, Foody JM, Gerber TC, Hinderliter AL, King SB, Kligfield PD, Krumholz HM, Kwong RYK, Lim MJ, Linderbaum JA, Mack MJ, Munger MA, Prager RL, Sabik JF, Shaw LJ, Sikkema JD, Smith CR, Smith SC, Spertus JA, Williams SV. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the diagnosis and management of patients with stable ischemic heart disease: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol* 2012;**60**:2564–2603.
24. Neglia D, Rovai D, Caselli C, Pietila M, Teresinska A, Aguade-Bruix S, Pizzi MN, Todiere G, Gimelli A, Schroeder S, Drosch T, Poddighe R, Casolo G, Anagnostopoulos C, Pugliese F, Rouzet F, Le Guludec D, Cappelli F, Valente S, Gensini GF, Zawaideh C, Capitanio S, Sambucetti G, Marsico F, Perrone Filardi P, Fernandez-Golfín C, Rincon LM, Graner FP, de Graaf MA, Fiechter M, Stehli J, Gaemperli O, Reyes E, Nkomo S, Maki M, Lorenzoni V, Turchetti G, Carpeggiani C, Marinelli M, Puzzuoli S, Mangione M, Marcheschi P, Mariani F, Giannessi D, Nekolla S, Lombardi M, Sicari R, Scholte AJ, Zamorano JL, Kaufmann PA, Underwood SR, Knuuti J; EVINCI Study Investigators. Detection of significant coronary artery disease by noninvasive anatomical and functional imaging. *Circ Cardiovasc Imaging* 2015;**8**:e002179.
25. Knuuti J, Ballo H, Juarez-Orozco LE, Saraste A, Kolh P, Rutjes AWS, Juni P, Windecker S, Bax JJ, Wijns W. The performance of non-invasive tests to rule-in and rule-out significant coronary artery stenosis in patients with stable angina: a meta-analysis focused on post-test disease probability. *Eur Heart J* 2018;**39**:3322–3330.
26. Katz DH, Deo RC, Aguilar FG, Selvaraj S, Martinez EE, Beussink-Nelson L, Kim KA, Peng J, Irvin MR, Tiwari H, Rao DC, Arnett DK, Shah SJ. Phenomapping for the identification of hypertensive patients with the myocardial substrate for heart failure with preserved ejection fraction. *J Cardiovasc Transl Res* 2017;**10**:275–284.
27. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, Berry J, Grodin JL, Pandey A. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail* 2020;**22**:148–158.
28. Shah SJ, Katz DH, Deo RC. Phenotypic spectrum of heart failure with preserved ejection fraction. *Heart Fail Clin* 2014;**10**:407–418.
29. Wang S, Khera R, Das SR, Vigen R, Wang T, Luo X, Lu R, Zhan X, Xiao G, Vongpatanasin W, Xie Y. Usefulness of a simple algorithm to identify hypertensive patients who benefit from intensive blood pressure lowering. *Am J Cardiol* 2018;**122**:248–254.
30. Cardiovascular Data Science (CarDS) Lab. ASSIST©. Available at: <https://www.cards-lab.org/assist> (accessed 8 April 2021).