Using machine learning to standardize diagnostic testing pathways in suspected coronary artery disease

By Dr. EK Oikonomou & Dr. R Khera

INTRODUCTION

Coronary artery disease (CAD) affects nearly 200 million people in the world, with chest pain representing one of the major presenting symptoms [1, 2]. Significant advances in the field of cardiovascular diagnostics have led to the development of several imaging and testing modalities, which can be used as gatekeepers prior to consideration of invasive coronary angiography. Such tests differ in their speed, costs, sensitivity, and specificity and their use is often guided by clinical reasoning as well as local expertise and availability [3].

Two main groups of diagnostic imaging modalities that are deployed in the investigation of chest pain are currently used in clinical practice [4]. Anatomical testing, through coronary computed tomography angiography (CCTA), enables the description of coronary anatomy and detection of structural abnormalities including luminal stenoses. On the other hand, functional testing, including stress electrocardiography, stress echocardiography, stress magnetic resonance imaging, and the most commonly used nuclear testing through single positron-emission computed tomography (SPECT) or positron emission tomography (PET) rely on the detection of regional ischemia through a combination of exercise/pharmacologic stimulation and diagnostic imaging [5].

For years, these were used interchangeably in the absence of clinical trial data comparing their efficacy and safety. However, in 2014, PROMISE (Prospective Multicenter Imaging Study for Evaluation of Chest Pain) [5] and in 2018, SCOT-HEART (Scottish COmputed Tomography of the HEART Trial) [6, 7] demonstrated that anatomical imaging has similar cardiovascular outcomes when compared to stress testing and may even improve long-term outcomes when used in addition to standard of care, including stress testing. To date, PROMISE remains the largest randomized controlled trial to have compared CCTA to functional testing in low-risk symptomatic patients with stable chest pain [5].

Since the publication of the PROMISE and SCOT-HEART trials, CCTA has gained ground as an alternative to functional imaging [4, 8]. However, the clinical equipoise on which test to select for each patient remains, with the most recent European Society of Cardiology (ESC) guidelines assigning Class I recommendation to both CCTA and non-invasive functional testing as appropriate initial tests to diagnose CAD in symptomatic patients [9].

PERSONALIZING THE INTERPRETATION OF CLINICAL TRIALS

In general, randomized trials assess the efficacy and safety of an intervention across a population, but in their standard form do not enable inference on the personalized benefit that each individual patient derives from an intervention A versus a second intervention B. To identify patient populations that derive differential benefit from either approach, subgroup analyses in PROMISE demonstrate 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]. However, such analyses dichotomize the phenotypic variation seen in the study population across a single axis and fail to account for large variation in demographic and clinical features within such subgroups.

In our work [13], we developed and validated a novel machine learning-based methodology that projects a trial's baseline population into a multidimensional space, where each dimension represents a phenotypic variable recorded prior to randomization, thus enabling a topological representation of the phenotypic variation seen in the study. In simple words, each individual is projected to a space, where their closest neighbours are characterized by a combination of phenotypic features spanning the full breadth of information recorded prior to randomization that most closely resembles that of the index patient. With increasing distance from each patient, the phenotypic similarity decreases. This enables the description of phenotypic neighbourhoods around each patient which included the 5% most similar study participants, thus providing a way of extracting individualized risk estimates of major adverse cardiovascular events (MACE) with anatomical versus functional testing. In a series of *in silico* experiments, each patient's neighbourhood formed the population for a simulated mini-trial, with the process repeated for each of the participants in the trial..

The Authors

Dr. EK Oikonomou[1] & Dr. R Khera[2,3]

1. Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

2. Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

3. Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA

Corresponding author: Dr. Rohan Khera,

Email: rohan.khera@yale.edu; @rohan_khera

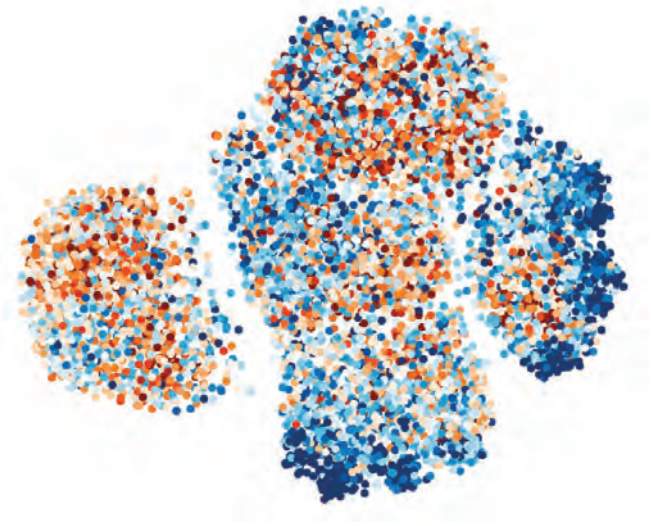

Figure 1. Risk phenomaps of the PROMISE trial. A manifold embedding of the baseline phenotypic variance seen in the PROMISE chest pain population based on 57 pre-randomization phenotypic traits. Labelling of the phenomaps with the neighbourhood-derived individualized risk estimates demonstrated distinct topological neighbourhoods favouring anatomical imaging or functional testing based on the observed risk in PROMISE. PROMISE: Prospective Multicenter Imaging Study for Evaluation of Chest Pain. (Reproduced with permission from Eur Heart J, ehab223, *https://doi.org/10.1093/eurheartj/ehab223*).

A MACHINE LEARNING METHOD TO PERSONALIZE THE INTERPRETATION OF CLINICAL TRIALS

Applying the method described above [13], we were able to uncover treatment effect heterogeneity across the PROMISE trial population, identifying phenotypic (topological) neighbourhoods where anatomical imaging was associated with a reduction in the risk of MACE compared to functional testing, and vice versa [Figure 1]. Whereas our analysis confirmed that across the study population, the two diagnostic strategies appear to be equivalent, it highlighted that at an individual level, certain patients benefit more from one strategy than the other.

GENERATION OF THE ASSIST TOOL

Having described the existence of treatment effect heterogeneity, we sought to identify which factors were most strongly associated with benefit from one strategy over the other, and built a parsimonious phenotypic signature that could be used prospectively to estimate the personalized benefit of anatomical versus functional testing. For that, we divided our PROMISE population randomly into an 80% subset, used for training and cross-validation, and 20% used for testing. We trained an extreme gradient boosting tree algorithm to predict the personalized relative hazard based on the topological analyses described above and used concepts from machine learning and game theory to assign relative feature importance values to each potential predictor. Using a combination of 15-features that most reliably and consistently correlated with personalized relative hazards favoring one strategy over another, we defined a decision support tool, named ASSIST©(Anatomical vs. Stress teSting decIsion Support Tool) [13]. To facilitate adoption of this tool for research purposes, we have made it available as part of an online browser-accessible online calculator (Cardiovascular Data Science (CarDS) Lab. ASSIST©: *https://www.cards-lab.org/assist*) [Figure 2] .

VALIDATION OF THE ASSIST TOOL

We validated ASSIST in the remaining 20% of PROMISE participants that were not included in its development as a part of an internal validation strategy. Furthermore, in a selected unmatched and propensity score-matched population of SCOT-HEART (external validation) we pursued external validation of ASSIST. Herein, we explictly accounted for the different design of SCOT-HEART compared to PROMISE. In SCOT-HEART, anatomical testing was added to standard of care, which in most patients also included stress electrocardiography, and therefore, we only included patients in the CCTA arm who underwent anatomical testing without antecedent stress test (anatomical-first arm), whereas in the standard care arm we included all individuals with an initial stress test (functional-first arm). We observed that in both the internal and external validation sets, agreement between the ASSIST recommendation and the actual test performed was associated with a significantly lower incidence of MACE, for both of PROMISE's and SCOT-HEART's primary endpoints [Figure 3]. Of note, a post hoc analysis of individual risk factors in the external validation cohort did not identify patients more likely to have favourable outcomes with anatomical vs. functional testing ($P_{interaction}$ = 0.79 for sex, 0.35 for hypertension, and 0.85 for diabetes mellitus), further highlighting the generalizability of our approach over broad subgroup assessments.

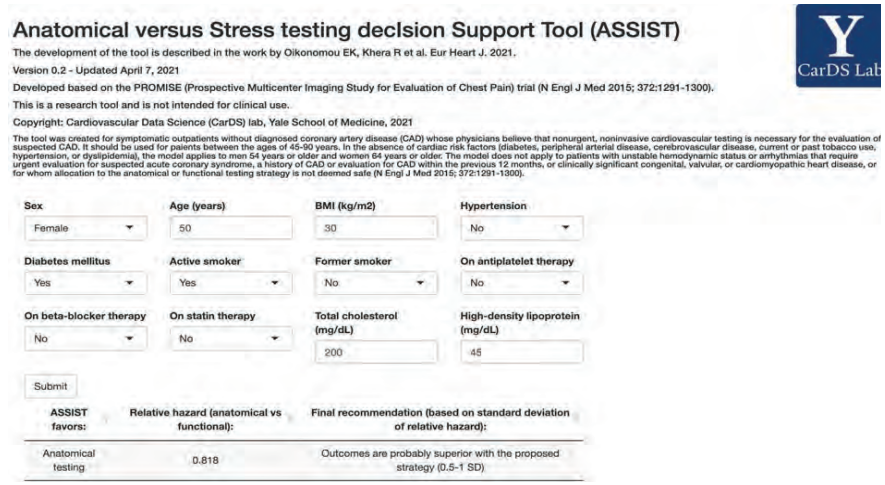

Figure 2. The ASSIST (Anatomical vs. Stress teSting decIsion Support Tool) tool. Available as part of an online browser-accessible online calculator (Cardiovascular Data Science (CarDS) Lab. ASSIST©: *https://www.cards-lab.org/assist (accessed* 11 June 2021)).

NEXT STEPS

In using the ASSIST calculator, one should be mindful of several considerations. First, this only applies to the patient population recruited as part of PROMISE, that means "*symptomatic outpatients without diagnosed CAD whose physicians believed that nonurgent, noninvasive cardiovascular testing was necessary for the evaluation of suspected CAD*". Notably, PROMISE investigators included participants with an age of more than 54 years (in men) or more than 64 years (in women) or an age of 45 to 54 years (in men) or 50 to 64 years (in women) with at least one cardiac risk factor (diabetes, peripheral arterial disease, cerebrovascular disease, current or past tobacco use, hypertension, or dyslipidemia).

Second, as described earlier, SCOT-HEART recruited a different population and instituted CCTA on top of standard care, which in most cases included stress testing. Our analysis in SCOT-HEART resulted in loss of randomization since we had to exclude patients whose management deviated from that of PROMISE, thus exposing these observations to potential confounding. However, the consistency of the internal and external validation was reassuring of our validity and generalizability.

Third, our analysis focused on cardiovascular endpoints, rather than diagnostic outcomes. Most of the literature has traditionally compared metrics of diagnostic accuracy and pre-, post-test probability using obstructive CAD on invasive coronary angiography as the gold standard [14]. Therefore, a discordance between diagnostic and hard clinical outcome measures should be accounted for when interpreting the results of our algorithm. Furthermore, since both PROMISE and SCOT-HEART were trials of diagnostic interventions, differences in outcomes are more likely to be explained by changes in therapeutic interventions and medications downstream of the diagnostic testing. Further research is needed to better understand these and explore to which extent they may be modifiable. Validation in prospective trials and real-world cohorts is needed and is currently underway.

CONCLUSIONS

In summary, we have recently developed an approach that defines an evidence-based strategy to pursue an anatomical versus functional evaluation of patients with suspected CAD.[13] In the first-ever application of a novel machine learning approach for personalized interpretation of clinical trial data, we were able to perform a series of local experiments to uncover and describe patterns of intervention effect heterogeneity in the PROMISE trial. A generalizable decision support tool derived from the PROMISE trial phenomap, named ASSIST [13], was validated in two geographically distinct large studies, and can be used to facilitate broader use of this information in shared decision-making in clinical practice.

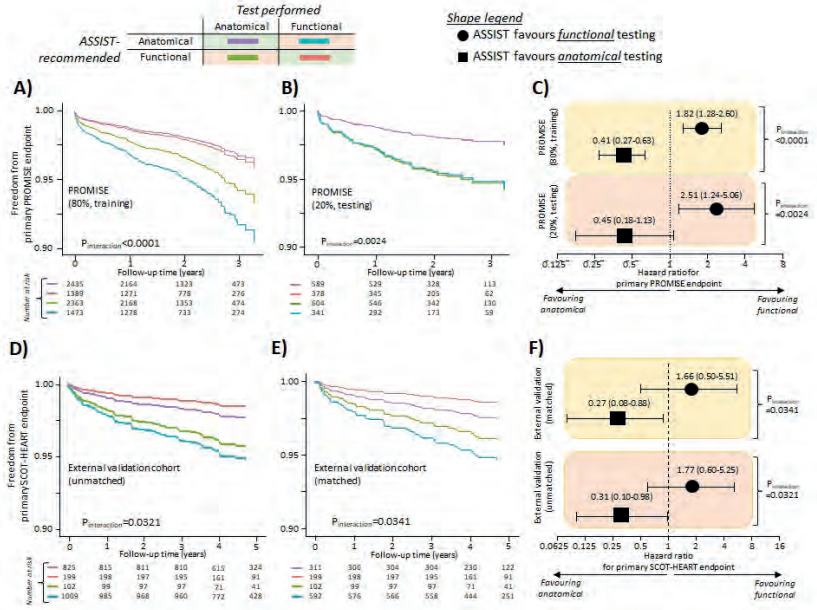

Figure 3. Internal and external validation of ASSIST. Application of the ASSIST tool in both the training and testing (internal validation) set of PROMISE demonstrated that concordance (vs. disagreement) between the ASSIST-proposed best initial diagnostic strategy and a patient random allocation to functional or anatomical testing was associated with an approximate two-fold reduction in the risk of the study primary composite endpoint (A-B: Adjusted survival curves; C: Forest Plot). These findings were replicated in the external validation cohort of SCOT-HEART for the primary composite endpoint of the study (D-E: Adjusted survival curves; F: Forest Plot). ASSIST, Anatomical vs. Stress teSting decIsion Support Tool; PROMISE, PROspective Multicenter Imaging Study for Evaluation of Chest Pain; SCOT-HEART, Scottish COmputed Tomography of the HEART Trial. (Reproduced and edited with permission from Eur Heart J, ehab223, https://doi.org/10.1093/eurheartj/ehab223).

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.

2. Davies SW. Clinical presentation and diagnosis of coronary artery disease: stable angina. Br Med Bull. 2001; 59: 17.

3. Poon M *et al.* Current Evidence and Recommendations for Coronary CTA First in Evaluation of Stable Coronary Artery Disease. J Am Coll Cardiol. 2020; 76: 1358.

4. Foy AJ *et al.* 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.

5. Douglas PS *et al.* Outcomes of anatomical versus functional testing for coronary artery disease. N Engl J Med. 2015; 372: 1291.

6. 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.

7. Scot-Heart Investigators *et al.* Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. N Engl J Med. 2018; 379: 924.

8. Moss AJ *et al.* 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 *et al.* 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J, 2020; 41: 407.

10. PagidipatiNJ *et al.* Sex Differences in Functional and CT Angiography Testing in Patients With Suspected Coronary Artery Disease. J Am Coll Cardiol. 2016; 67: 2607.

11. Sharma A *et al.* Stress Testing Versus CT Angiography in Patients With Diabetes and Suspected Coronary Artery Disease. J Am Coll Cardiol. 2019; 73: 893.

12. Hemal K *et al.* 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.

13. Oikonomou EK *et al.* A phenomapping-derived tool to personalize the selection of anatomical vs. functional testing in evaluating chest pain (ASSIST). Eur Heart J. 2021 ehab223. doi: 10.1093/eurheartj/ehab223.

14. Montalescot G *et al.* 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology. Eur Heart J. 2013; 34: p. 2949.