

BMJ Open Cohort profile: the Emory Cardiovascular Biobank (EmCAB)

Yi-An Ko,¹ Salim Hayek,² Pratik Sandesara,² Ayman Samman Tahhan,² Arshed Quyyumi²

To cite: Ko Y-A, Hayek S, Sandesara P, *et al.* Cohort profile: the Emory Cardiovascular Biobank (EmCAB). *BMJ Open* 2017;7:e018753. doi:10.1136/bmjopen-2017-018753

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-018753>).

Received 20 July 2017
Revised 16 October 2017
Accepted 18 October 2017

ABSTRACT

Purpose The Emory Cardiovascular Biobank (EmCAB) is an ongoing prospective registry of patients undergoing cardiac catheterisation, which was established to identify novel factors associated with the pathobiological process and treatment of cardiovascular disease.

Participants Individuals aged 18 years and older undergoing cardiac catheterisation at three Emory Healthcare sites in Atlanta are asked to participate in this prospective registry. Around 95% agree to participate. Around 7000 unique patients have been enrolled. The current data set contains detailed phenotyping, patient outcomes, genomics, protein biomarkers, regenerative markers, transcriptomic analysis, metabolomics profiling and longitudinal follow-up for adverse cardiovascular outcomes.

Findings to date Thus far, the EmCAB has approximately 3000 major cardiovascular events. About 48% of the EmCAB participants have more than 5 years of follow-up. It is a great resource for discovery of novel predictive factors for cardiovascular disease outcomes, including genomics, transcriptomics, protein biomarkers, oxidative stress markers and circulating progenitor cells. Several circulating inflammatory markers have shown to improve risk prediction metrics beyond standard risk factors.

Future plans Future integrative –omics analyses will provide the cardiovascular research community opportunities for subsequent mechanistic confirmation studies, which will promote the development of effective personalised therapy that leads to clinical care tailored to the individual patient.

INTRODUCTION

Cardiovascular disease (CVD) is the most common underlying cause of death in the world, accounting for an estimated 31% of all global deaths. By 2030, 44% of the US adult population is projected to have some form of CVD.¹ Although death rates attributable to CVD declined 40% between 1980 and 2000² and 25% from 2004 to 2014¹ in the USA, the incidence and prevalence continue to increase.³ In addition, CVD accounts for a significant burden of morbidity and health expenditures in the USA.⁴ As such, a public health priority is to develop precise risk assessment tools and cost-effective prevention and therapeutic strategies. While traditional

Strengths and limitations of this study

- The Emory Cardiovascular Biobank is enriched for patients with coronary artery disease with high event rates for major cardiovascular outcomes, which provides ample statistical power for studies of secondary risk prediction and prevention strategies.
- The availability of multiple –omics data (eg, genomics, transcriptomics, metabolomics) provides opportunities to discover novel markers for outcome prediction in the context of precision medicine.
- Patients are recruited from a single healthcare system. Although treatment heterogeneity is minimised, the results might not be generalisable to other patients.
- Depending on available resources, not all data types are available among all the patients.

risk factors for CVD (comprising age, sex, race, hypertension, dyslipidaemia, smoking and diabetes) predominantly account for the risk of *developing* coronary artery disease (CAD),^{5 6} they are less useful in predicting patient outcomes among those who already have CAD (c-statistic for 10-year CVD risk ≈ 0.82 vs c-statistic for secondary cardiovascular events ≈ 0.62).^{7 8} Novel biomarkers and genetic profile may better address the deficiencies in risk assessment. There is an urgent and compelling need for identifying novel biomarkers and for developing precise diagnostic tests to better risk stratify patients with established CAD in order to improve disease management strategies.

CAD has a complex multifactorial aetiology. The development of coronary atherosclerosis is a consequence of genetic susceptibility, environmental exposures and interactions between risk factors on many levels. Susceptibility to CAD varies extensively between subjects and may be partly attributed to genomic and epigenetic variations. Moreover, environmental, behavioural and psychosocial factors have a substantial impact on the incidence and progression of CAD. Importantly, atherosclerosis can be viewed as a progressive disease with a subclinical



CrossMark

¹Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

²Division of Cardiology, Emory University School of Medicine, Atlanta, Georgia, USA

Correspondence to
Dr Arshed Quyyumi;
aquyyum@emory.edu

phase that may precede development of plaque and is manifested as vascular dysfunction. Once atherosclerotic plaque formation occurs, the disease may adopt a long stable period when the plaque may grow slowly without any symptoms or with stable ischaemic symptoms. Alternatively, CAD may progress to unstable or acute coronary syndromes, including unstable angina, myocardial infarction (MI) and even sudden cardiac death. The underlying pathophysiological triggers and mechanisms for each of these phases are distinctly different. These factors are likely to be discovered only by careful characterisation of these phenotypes and subsequent application of gene sequencing, gene expression and epigenetic analyses, and ultimately by the study of proteins and metabolites.

The Emory Cardiovascular Biobank (EmCAB) was established in 2003 for detailed phenotyping and follow-up of a large population with and without obstructive CAD. The goal was to identify novel factors associated with the pathobiological process and treatment of CVD. The EmCAB is a prospective registry of patients undergoing cardiac catheterisation at three Emory Healthcare sites: Emory University Hospital, Grady Memorial Hospital and Emory Midtown hospital, supported by the Woodruff Foundation at Emory University, several NIH and other grants. Over the years, the size of EmCAB continued to grow with 7000 unique patients recruited to date. The current EmCAB data set contains full information on clinical variables including detailed family history and psychosocial factors obtained by questionnaires and data extracted from electronic health records on the entire population. In various subsets of the dataset, we have vascular function measures, outcomes, genomics, protein biomarkers, regenerative markers, transcriptomic analysis, metabolomics profiling and longitudinal follow-up for adverse cardiovascular outcomes. The rich and complex data available in EmCAB provide unparalleled opportunities for conduct of integrative analysis of various important, unsolved research questions, such as -omic signatures for CAD severity, novel predictors for disease progression and for incident cardiovascular events.

COHORT DESCRIPTION

All patients aged 18 years and older undergoing cardiac catheterisation at three Emory Healthcare sites in Atlanta are asked to participate in this prospective registry. Patients are excluded if they have congenital heart disease, severe valvular heart disease, severe anaemia, a recent blood transfusion, myocarditis, history of active inflammatory disease, cancer or are unable or not willing to provide consent (approximately 5%). The study was approved by the institutional review board (IRB) at Emory University (Atlanta, Georgia, USA) and is renewed annually. All participants provided written informed consent at the time of enrolment. Specimens and data continue to be added to the EmCAB as new subjects are consented on an almost daily basis. To facilitate future research, the

Table 1 Baseline characteristics of the Emory Cardiovascular Biobank cohort

| Variable | Female (n=2486) | Male (n=4371) |
|--------------------------|-----------------|---------------|
| Age (years) | 63.2±22.3 | 62.7±12.3 |
| Race | | |
| Caucasian white | 1634 (65.7) | 3291 (75.3) |
| African American black | 769 (30.9) | 857 (19.6) |
| Hispanic | 24 (1.0) | 48 (1.1) |
| Asian | 27 (1.1) | 86 (2.0) |
| Other | 32 (1.3) | 89 (2.0) |
| Marital status | | |
| Married | 1290 (51.9) | 3308 (75.7) |
| Widowed | 500 (20.1) | 228 (5.2) |
| Divorced | 418 (16.8) | 430 (9.8) |
| Separated | 62 (2.5) | 73 (1.7) |
| Never married | 216 (8.7) | 332 (7.6) |
| Employment status | | |
| Employed | 694 (27.9) | 1659 (38.0) |
| Unemployed | 249 (10.0) | 198 (4.5) |
| Disabled | 452 (18.2) | 617 (14.1) |
| Retired | 1091 (43.9) | 1897 (43.4) |
| Education level | | |
| Less than high school | 399 (16.05) | 521 (11.92) |
| High school graduate | 767 (30.85) | 1177 (26.93) |
| Some college | 622 (25.02) | 957 (21.89) |
| College degree and above | 698 (28.08) | 1716 (39.26) |
| History of smoking | 1384 (55.67) | 2955 (67.6) |
| Hypertension | 1920 (79.97) | 3330 (78.1) |
| Diabetes | 838 (34.97) | 1509 (35.45) |
| Dyslipidaemia | 1638 (68.22) | 3081 (72.32) |
| Chronic kidney disease | 180 (8.75) | 413 (11.33) |
| History of MI | 469 (19.67) | 1084 (25.67) |
| History of heart failure | 902 (36.28) | 1614 (36.93) |
| Obstructive CAD | 1417 (68.75) | 3086 (81.04) |

Mean±SD is presented for age, and n (%) is presented for all other variables.

CAD, coronary artery disease; MI, myocardial infarction.

biobank has IRB approval for future testing of biospecimens and analysis of de-identified data.

Table 1 summarises the patient characteristics in this cohort. The mean age at enrolment was 63. Approximately two thirds were male, 72% white, 24% black, 2% Asian and 1% Hispanic. While 75% of them had obstructive coronary artery disease (defined as >50% luminal narrowing in a major epicardial vessel) at enrolment, 11% had normal (<10% stenosis) coronary angiograms. About 25% had a history of previous MI, 8% presented with an acute MI and 36% had a history of heart failure. Most

(44%) were retired, 34% were employed, 16% disabled and 6% unemployed. About 51% were high school graduates with/without some college education, and 36% had a college degree (16% received graduate education). The majority (67%) were married; 25% were widowed, divorced or separated; and 8% were never married.

Follow-up measures

Follow-up phone interview is conducted by trained study personnel, and the calls are planned to be made at 1 and 5 years after initial enrolment. The actual follow-up time periods range up to 12 years because of limited resources. Adverse CVD events, including death, non-fatal MI, acute coronary syndromes, heart failure, hospitalisations, cardiac procedures (eg, revascularisation), strokes and peripheral arterial disease events are recorded. If a patient has experienced a major CVD event such as cardiovascular death, MI, hospitalisations or stroke, medical records are used for verification. When a study participant has died, the study personnel gather information from multiple sources to confirm the event and the time of the event. Social security death index, state records, obituaries, hospital records and death certificates are queried, and family members are contacted to obtain details. An Events Adjudication Committee designates final arbitration of all cardiovascular events and adjudicates the cause of death.

Additional information is collected on follow-up coronary angiogram data, development of comorbidities (hypertension, diabetes, cardiac arrhythmia, valvular heart disease, obstructive sleep apnoea and cancer) and changes in medications (aspirin, Plavix, beta blocker, calcium channel blocker, angiotensin-converting-enzyme inhibitors/angiotensin receptor blockers (ACEI/ARB), aldosterone antagonist, diuretics, nitrates, lipid-lowering agents, anticoagulants). Furthermore, detailed questionnaires are collected to assess for depression (Patient Health Questionnaire 9), angina frequency (Seattle Angina Questionnaire), physical activity (Duke Activity Status Index), functional status (New York Heart Association classification) and smoking status.

Major adverse event rate (for all-cause death or MI) is 6.4% in 1 year and 13% in 3 years, and approximately 1500 major events have been recorded. All event rates including revascularisation are 17% in 1 year and 25% in 3 years, for a total of approximately 3000 events to date. **Figure 1** displays the cumulative incidence over time for each of the major cardiovascular events. Approximately 96% of patients have follow-up data for a median follow-up time of 3.5 years (IQR: 1.7–6.7 years), with a 3.3% refusal rate for continued follow-up. These patients tend to be healthier (hypertension 72% vs 79%, $P=0.01$; diabetes 29% vs 35%, $P=0.05$; dyslipidaemia 65% vs 71%, $P=0.04$; obstructive CAD 70% vs 77%, $P=0.02$; history of heart failure 29% vs 37%, $P=0.02$), compared with the group who agree to be contacted. Also, more of these patients have full-time jobs (36% vs 29%, $P=0.01$).

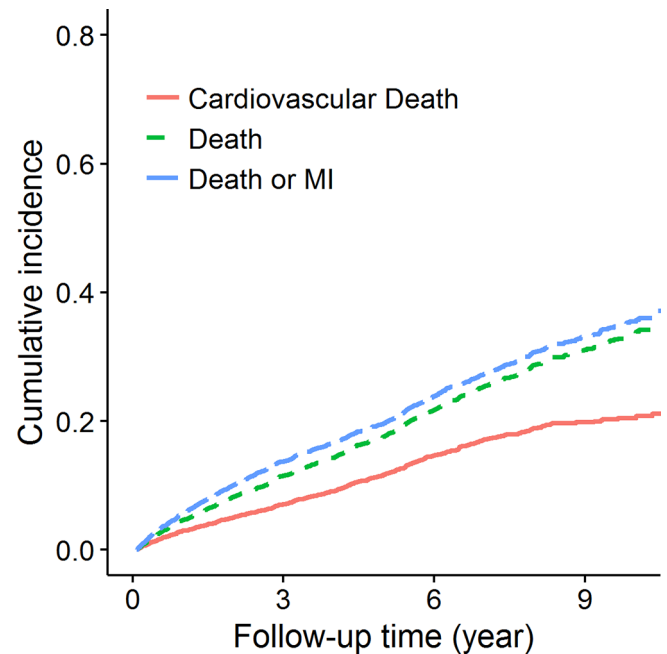


Figure 1 Cumulative incidence for cardiovascular death, all-cause death, and death/MI in the Emory Cardiovascular Biobank. MI, myocardial infarction.

Study measures

At enrolment, patients are interviewed to collect information on demographic characteristics, medical history, detailed family history, medication usage, health behaviours (alcohol/drug use), sleep quality, menstrual history, psychological factors and neuropsychological functioning prior to cardiac catheterisation. In addition, medical records are reviewed by study personnel to confirm self-reported history of MI and other chronic conditions as well as to document previous angiographic findings and coronary revascularisation history. Laboratory data collection includes sodium, potassium, creatinine, albumin, B-type natriuretic peptide, blood urea nitrogen, haemoglobin, glucose, HbA1c (glycated haemoglobin), complete blood count, serum lipid profile. Comprehensive angiographic results at the time of enrolment are recorded. Furthermore, subjects are followed at 1 and 5 years. To date, detailed follow-up data (as described above) are available from 96% of the patients.

Additional blood samples used to obtain biochemical data (ie, DNA, RNA, progenitor cells, proteomics, metabolomics and biomarker assays) are collected by a nurse or physician from the arterial catheter inserted for cardiac catheterisation before administration of anticoagulants after an overnight fasting, and stored in Paxgene tubes (QIAGEN, San Diego, California, USA) at -80°C . These data are available in subsets of patients; specifically, we have conducted high-resolution metabolomics analyses for >20 000 metabolites using liquid chromatography mass spectrometry (LC/MS) in 1500 EmCAB patients. Samples are analysed in triplicate by liquid chromatography–Fourier transform mass spectrometry (Accela-LTQ Velos Orbitrap; m/z range from 85 to 850) with $10\mu\text{L}$

Table 2 Summary of data collected in the Emory Cardiovascular Biobank

| Phase | Category | Measurements |
|--------------------|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Baseline enrolment | Demographics/ socioeconomics | Age, sex, race, education, marital status, employment status, zip code |
| | Questionnaires | Patient Health Questionnaire 9 for depressive symptoms, ²⁵ stress questionnaire, Pittsburgh Sleep Quality Index, ²⁶ health behaviours (smoking and alcohol consumption, Duke Activity Status Index ²⁷), medical history, family history of disease, menstrual history, Seattle Angina Questionnaire for angina symptoms ^{28,29} |
| | Admission details | Height, weight, blood pressure, heart rate, vascular functions and ECG, medications |
| | Chart review | Stress test results, angiogram, history and details of MI, CABG, PCI, stroke, heart failure, transplant, valve disease and other chronic conditions |
| | Laboratory readings | Sodium, potassium, creatinine, albumin, B-type natriuretic peptide, blood urea nitrogen, haemoglobin, glucose, HbA1c, complete blood count, serum cholesterol and fractioned lipid profile |
| | Angiogram | History of coronary artery bypass, bypass grafts, stenosis, intervention |
| | Genetics | GWAS data (on approximately 1000 patients) and candidate SNPs (on approximately 3000 patients) |
| | Transcriptomics | 14343 probes consistently detected above background in multiple gene expression datasets (on 338 patients) |
| | Circulating progenitor cells | CD45med+ cells with and without co-expression of CD133 (CD34+/CD133+), VEGFR2 (CD34+/VEGFR2 and CD34+/CD133+/VEGFR2+) and CXCR4 (CD34+/CXCR4+ and CD34+/CD133+/CXCR4+) on approximately 2200 patients |
| | Metabolomics | Three readings per subject on approximately 1500 patients |
| Follow-up | Protein biomarkers | Hs-cTn, hs-CRP, FDP, HSP70, suPAR, CMV, PCSK9 |
| | Oxidative stress markers | CyS, CySS, GSH and GSSG |
| | Adverse events | Death, MI, heart failure, revascularisation, stroke |

CABG, coronary artery bypass graft surgery; CMV, cytomegalovirus; CyS, cysteine; CySS, cystine; FDP, fibrin degradation product; GSH, reduced glutathione; GSSG, oxidised glutathione; GWAS, genome wide association study; HbA1c, haemoglobin A1c; hs-CRP, high sensitivity C reactive protein; Hs-cTn, high sensitivity cardiac troponin; HSP70, heat shock protein-70; MI, myocardial infarction; PCI, percutaneous coronary intervention; PCSK9, proprotein convertase subtilisin/kexin type 9; SNP, single nucleotide polymorphism; suPAR, soluble urokinase-type plasminogen activator receptor.

injection volume using a dual chromatography setup (anion exchange and C18) and a formic acid/acetonitrile gradient. Extractions are performed with acetonitrile containing a mixture of internal standards and maintained in an autosampler maintained at 4°C until injection. Electrospray ionisation is used in the positive ion mode. Data are extracted using apLCMS⁹ with modifications by xMSanalyzer¹⁰ as m/z features, where an m/z feature is defined by m/z (mass-to-charge ratio), retention time and ion intensity (integrated ion intensity for the chromatographic peak). Identities of many of the m/z features are known from previous research using ion dissociation patterns by tandem mass spectrometry, coelution with authentic standards and cross-platform validation. Possible identities of other m/z features were obtained using the Metlin Mass Spectrometry Database and Kyoto Encyclopedia of Genes and Genomes (KEGG) (<http://www.genome.jp/kegg/pathway.html>) databases using a mass error threshold of 10 ppm (relative m/z error ×106).

A total of 105 candidate single nucleotide polymorphisms (SNPs) were analysed in a subset of up to 3488 participants and genome-wide genotyping (Illumina

Multi-Ethnic Global array) in approximately 1300 patients. Gene expression data (Illumina HT-12 bead arrays) with 14111 probes are available in 338 patients. We have performed extensive phenotyping of circulating progenitor cells (CD34+ cell subsets), a measure of endogenous regenerative capacity in >2000 patients. The availability of immunoassays and high-performance LC/MS has enabled the identification of inflammatory protein biomarkers. Circulating levels of the following protein biomarkers were measured in 3500–6000 participants: hs-cTn (high-sensitivity cardiac troponin), hs-CRP (high-sensitivity C reactive protein), suPAR (soluble urokinase-type plasminogen activator receptor), FDP (fibrin degradation product) and HSP-70 (heat shock proteins). High-sensitivity troponin-I was measured by Abbott Laboratories (Abbott Park, Illinois, USA). Serum CRP and FDP measurements are determined using a sandwich immunoassay by FirstMark, Division of GenWay Biotech (San Diego). Serum HSP-70 is measured with a sandwich ELISA (R&D Systems, Minneapolis, Minnesota, USA) and optimised by FirstMark. Plasma suPAR levels are measured using

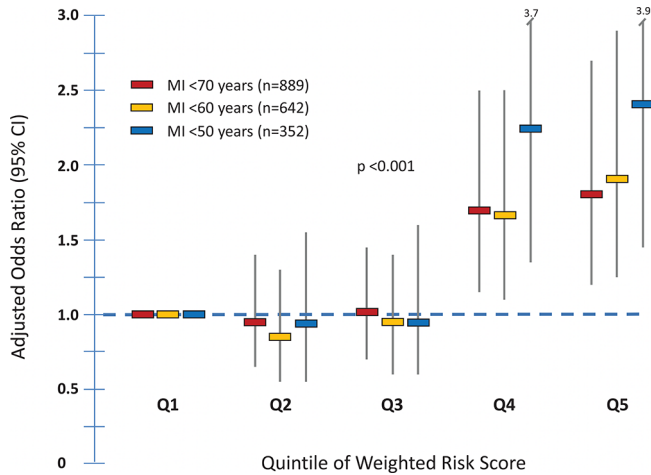


Figure 2 Association between five quintile groups of genetic risk score and myocardial infarction (MI) in different age groups within the Emory Cardiovascular BioBank. Adjusted ORs and 95% CIs are presented. Covariates included sex, body mass index, diabetes, hypertension, hyperlipidaemia, smoking and family history of coronary artery disease.

commercially available kits (suPARnostic kit; Virogates, Copenhagen, Denmark). Oxidative markers including plasma amino thiols, cysteine and glutathione and their oxidised counterparts are available in 1500 patients. Table 2 summarises the types of patient data available in the EmCAB database.

In addition to the data listed in table 2, Emory Clinical Data Warehouse provides abstraction of comprehensive and longitudinal electronic health record data of clinical characteristics, comorbidities and laboratory values (eg, blood pressure, body weight, HbA1c, urine protein and serum creatinine). Abstraction is possible for all visits before and after the enrolment date at Emory. Moreover, abstraction can be customised to obtain data in a specific time period depending on the research purposes. In summary, the large sample size combined with a high event rate in this cohort with CAD provides a unique opportunity to conduct CVD outcome studies to answer a variety of research questions that may not be possible in other population-based cohorts.

FINDINGS

The EmCAB has been used as a data resource for discovery of novel predictive factors (including biomarkers and molecular signatures) for clinical outcomes. *Genomics*: We demonstrated that the 9p21 risk locus promotes atherosclerosis and its progression in 2334 Caucasian patients enrolled in EmCAB.¹¹ Moreover, in 308 EmCAB patients who underwent two or more coronary angiograms at least 6 months apart, we measured the change in Gensini score over time and found an association between the rs10757278 SNP and CAD progression ($P=0.023$) with homozygotes for the risk variant having >3-fold greater odds of CAD progression compared with the wild-type group. Subsequently, a combined genetic risk score (GRS) based on 11 disease-associated SNPs was developed to predict adverse cardiovascular events. In 2597 EmCAB subjects, we identified cases with a history of MI onset at age <50 years and controls ≥ 50 years old without prior MI. The analysis was repeated with 60 and 70 years old as cut points. The GRS was found to be significantly associated with prevalent MI (figure 2) and improved the c-statistic by up to 1.7%.¹² In addition, the EmCAB has involved in many large-scale genetic association studies, including finding variants for atherosclerosis,^{13 14} CAD,^{15–17} MI¹⁸ and mortality.^{13 14 19} In recent years, the EmCAB has also participated in the GENIUS-CHD consortium, a global collaborative effort to evaluate the effect of genetic variation and identify determinants for subsequent coronary heart disease outcomes. *Transcriptomics*: We further explored peripheral blood gene expression profiles to assess its ability to predict acute MI and/or cardiovascular death. In 338 EmCAB subjects with CAD, we analysed gene expression using the Illumina HT-12 microarrays in two phases (discovery $n=175$ and replication $n=163$) and genome-wide association study was performed. There was significant differential expression between healthy and MI groups with overall downregulation of genes involved in T-lymphocyte signalling and upregulation of inflammatory genes. Expression quantitative trait loci analysis provided evidence for altered local genetic regulation of transcript abundance in MI samples. A first principal component (PC1) score derived from 10 transcripts that capture the covariance of 238 genes in the discovery

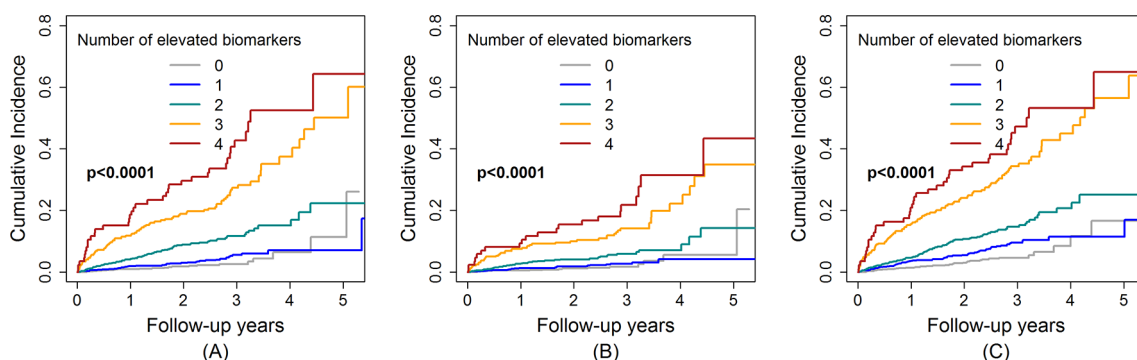


Figure 3 Cumulative incidence plots for all-cause death (A), cardiovascular death (B) and all-cause death/myocardial infarction (C) per category of the biomarker risk score.

phase significantly predicted risk of cardiovascular death in the replication and combined samples. The hazard was significantly higher for those with a low PCI score compared with those with a high score and improved the c-statistic by 9% ($P=0.03$).²⁰ **Protein biomarkers:** The availability of immunoassays has enabled the identification of inflammatory protein biomarkers CRP, FDP, HSP-70 and suPAR as prognostic predictors of future risk of death and MI in patients with CAD. We combined these biomarkers into an aggregate biomarker risk score, calculated by adding one point for every elevated marker that stratifies the risk of major adverse events in approximately 3000 EmCAB patients (figure 3).^{21 22} The results suggest the importance of activation of multiple pathophysiological pathways, measured in the circulation, as directly impacting prognosis. **Oxidative stress:** We have also shown that a high burden of oxidative stress, measured as higher levels of cystine, lower levels of glutathione or altered ratios of oxidised to reduced aminothiols, is associated with cellular dysfunction, ageing, risk factors for CVD and with incident and MI in the 1411 EmCAB patients with CAD.²³ **Circulating progenitor cells:** Reduced circulating progenitor cell counts (CD34+ or CD34+/CD133+), indicating decreased reparative capacity, were found to be related to an elevated risk of death in these patients.²⁴ Importantly, CD34+/CD133+ cell counts improved risk prediction metrics beyond standard risk factors.

STRENGTHS AND WEAKNESSES

The goal of the EmCAB is to identify novel determinants of CVD progression and outcomes. The main advantages of this cohort include extensive collection of epidemiological and detailed clinical data, a large patient cohort from a single healthcare system that minimises treatment heterogeneity, and long-term follow-up information of patients' clinical data and outcomes. Most importantly, this patient population has coronary angiographic information on all enrolled subjects, is enriched for presence of CAD, and has high event rates for various major cardiovascular outcomes that provide ample statistical power for studies of secondary risk prediction and prevention strategies. Furthermore, the availability of genetic and molecular profiling data provides opportunities to discover important features to improve prediction of clinical outcomes in the context of precision medicine. A potential weakness is that patients are only recruited from Emory Healthcare system and the results may not be generalisable to other patients with CAD. Another limitation is that available resources evolve over time, so not all data types are available among all patients. With the ultimate goal of recruiting a total of 12 000 EmCAB patients, the sample size for each data type continues to increase.

Acknowledgements We are extremely grateful for our participants' support. We would like to acknowledge the members of the Emory Biobank Team, Emory Clinical Cardiovascular Research Institute and Atlanta Clinical and Translational Science Institute for the recruitment of participants, compilation of data and preparation of samples. Specifically, we would like to thank M Awad, A Alkholder, J

Hartsfield, N Ghasemzadeh, D Eapen, R Patel, H Aida, M Gafeer and N Abdelhadi for assisting in data collection.

Contributors YK manages the EmCAB data and wrote the manuscript. SSH re-designed the database and updated the data collection tools, which substantially improved the data quality. AAQ is the PI of the EmCAB and established the Emory Biobank. SSH, PS, AST and AAQ revised the manuscript critically. All authors agree to be accountable for all aspects of this work.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Emory IRB.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The EmCAB committee welcomes collaboration and the database will be shared with investigators. External researchers can get access to the EmCAB data through contacting and obtaining an authorisation by the Principal Investigator Arshed Quyyumi (aquyyum@emory.edu).

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation* 2017;135:e146-03.
2. Ford ES, Ajani UA, Croft JB, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. *N Engl J Med* 2007;356:2388-98.
3. Davies AR, Smeeth L, Grundy EM. Contribution of changes in incidence and mortality to trends in the prevalence of coronary heart disease in the UK: 1996-2005. *Eur Heart J* 2007;28:2142-7.
4. Trogon JG, Finkelstein EA, Nwaise IA, et al. The economic burden of chronic cardiovascular disease for major insurers. *Health Promot Pract* 2007;8:234-42.
5. Kannel WB, Dawber TR, Kagan A, et al. Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study. *Ann Intern Med* 1961;55:33-50.
6. D'Agostino RB, Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743-53.
7. Ferret BS, van Kempen BJ, Hunin MG, et al. Predictive value of updating Framingham risk scores with novel risk markers in the U.S. general population. *PLoS One* 2014;9:e88312.
8. Zengin E, Bickel C, Schnabel RB, et al. Risk factors of coronary artery disease in secondary prevention-results from the atherogenesis study. *PLoS One* 2015;10:e0131434.
9. Yu T, Park Y, Johnson JM, et al. aPLCMS--adaptive processing of high-resolution LC/MS data. *Bioinformatics* 2009;25:1930-6.
10. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics* 2013;14:15.
11. Patel RS, Su S, Neeland IJ, et al. The chromosome 9p21 risk locus is associated with angiographic severity and progression of coronary artery disease. *Eur Heart J* 2010;31:3017-23.
12. Patel RS, Sun YV, Hartiala J, et al. Association of a genetic risk score with prevalent and incident myocardial infarction in subjects undergoing coronary angiography. *Circ Cardiovasc Genet* 2012;5:441-9.
13. Reilly MP, Li M, He J, et al. Identification of ADAMT5 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* 2011;377:383-92.
14. Helgadottir A, Gretarsdottir S, Thorleifsson G, et al. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. *J Am Coll Cardiol* 2012;60:722-9.

15. Schunkert H, König IR, Kathiresan S, *et al*. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 2011;43:333–8.
16. Chan K, Patel RS, Newcombe P, *et al*. Association between the chromosome 9p21 locus and angiographic coronary artery disease burden: a collaborative meta-analysis. *J Am Coll Cardiol* 2013;61:957–70.
17. Björnsson E, Gudbjartsson DF, Helgadóttir A, *et al*. Common sequence variants associated with coronary artery disease correlate with the extent of coronary atherosclerosis. *Arterioscler Thromb Vasc Biol* 2015;35:1526–31.
18. Gretarsdóttir S, Baas AF, Thorleifsson G, *et al*. Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. *Nat Genet* 2010;42:692–7.
19. Morgan TM, House JA, Cresci S, *et al*. Investigation of 95 variants identified in a genome-wide study for association with mortality after acute coronary syndrome. *BMC Med Genet* 2011;12:127.
20. Kim J, Ghasemzadeh N, Eapen DJ, *et al*. Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death. *Genome Med* 2014;6:40.
21. Eapen DJ, Manocha P, Patel RS, *et al*. Aggregate risk score based on markers of inflammation, cell stress, and coagulation is an independent predictor of adverse cardiovascular outcomes. *J Am Coll Cardiol* 2013;62:329–37.
22. Ghasemzadeh N, Hayek SS, Ko YA, *et al*. Pathway-specific aggregate biomarker risk score is associated with burden of coronary artery disease and predicts near-term risk of myocardial infarction and death. *Circ Cardiovasc Qual Outcomes* 2017;10:e001493.
23. Patel RS, Ghasemzadeh N, Eapen DJ, *et al*. Novel biomarker of oxidative stress is associated with risk of death in patients with coronary artery disease. *Circulation* 2016;133:361–9.
24. Patel RS, Li Q, Ghasemzadeh N, *et al*. Circulating CD34+ progenitor cells and risk of mortality in a population with coronary artery disease. *Circ Res* 2015;116:289–97.
25. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
26. Buysse DJ, Reynolds CF, Monk TH, *et al*. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193–213.
27. Hlatky MA, Boineau RE, Higginbotham MB, *et al*. A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index). *Am J Cardiol* 1989;64:651–4.
28. Spertus JA, Winder JA, Dewhurst TA, *et al*. Monitoring the quality of life in patients with coronary artery disease. *Am J Cardiol* 1994;74:1240–4.
29. Spertus JA, Winder JA, Dewhurst TA, *et al*. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. *J Am Coll Cardiol* 1995;25:333–41.