The GENVASC Study

Genetics and the Vascular Health Check Programme

Chief Investigator: Professor Nilesh J Samani

Co-investigators: Professor Azhar Farooqi
Professor Kamlesh Khunti
Dr Nitin Joshi
Professor Mayur Lakhani
Dr Shora Montgomerie
Dr Azar Zafar
Dr Amardeep Heer

Statistical adviser: Professor John Thompson

Background

Coronary artery disease (CAD) is the commonest cause of premature death and disability in the UK. Several demographic and lifestyle factors, such as age, gender, smoking, hypertension, diabetes and dyslipidemia contribute to risk of CAD. A number of CAD prediction risk algorithms based on these factors, such as the Framingham and the QRisk2 scores have been developed and allow classification of individuals into low (< 10%), medium (10-20%) and high (> 20%) 10-year CAD risk. These risk scores have been used to identify and target primary prevention measures to those at highest risk. While targeting such individuals is clearly beneficial, because many more subjects are located in the intermediate or low risk categories, although their proportional risk is lower, in absolute terms more events actually occur in these groups. ***Improving the accuracy of risk categorisation for CAD is therefore a high public health and clinical priority.***

Inheritance plays an important role in the aetiology of CAD. The risk to an individual is 4-8 fold higher if a first degree relative has died prematurely of CAD. The heritability of CAD is estimated at around 50%. In some, especially more recent risk scores, a "family history" of CAD is included in the algorithm. However, identifying a positive family history due to inheritance has significant limitations. Family history based on recall can be notoriously inaccurate. Algorithms vary in the age cut-off used to define a positive family history. Furthermore, an individual's family may not be sufficiently large (e.g, no siblings) to assess familial risk, family members could have died from competing causes (e.g. cancer or road traffic accidents) before manifesting CAD, or could have developed CAD but due to a strong lifestyle factor such as heavy smoking. ***In short, although a family history of CAD can be useful it is neither a sufficient or accurate surrogate for an individual's genetic risk.***

Recently, significant progress has been made in directly dissecting the genetic basis of CAD. In work led by the principal investigator in collaboration with national and international collaborators, over 30 common genetic variants (carried by between 10-80% of the population) have been identified that increase risk of CAD by between 5-30% per copy of each variant. Further variants are likely to emerge in on-going work and especially lower frequency variants that have more powerful effects. Individually, the genetic variants do not have sufficient discrimination to individually change risk prediction sufficiently. However, a Genetic Risk Score (GRS) based on combining the variants (adjusted for their individual effects) could be more powerful. Indeed, in a recent study we showed that there was a > 3-fold difference in odds ratio for CAD between those subjects in the highest quintile compared with those in the lowest quintile for a GRS score based on 25 of the initially identified CAD-associated variants. This is similar to or greater than the strength of association seen with other established risk factors such as blood pressure and cholesterol. Addition of further variants as they are discovered to the GRS is likely to further improve its risk prediction potential. ***Therefore, recent discoveries on the basis of CAD now provide a framework for testing whether adding genetic information in the form of a genetic risk score can improve current risk prediction of CAD.***

To test whether a GRS for CAD can improve risk prediction requires assembling a large cohort of individuals representative of the general population who are: (i) free of overt CAD at recruitment (ii) who are assessed in a uniform fashion for their CAD risk (iii) and who can provide blood samples for genetic analysis and (iv) who can be followed up systematically for CVD outcomes. Such cohorts are rare nationally and internationally. Biobank UK (at a cost of > £30 million) has been established to address this deficiency although one of its limitations is that it is not representative of the general population. One of the key factors that inhibits the assembly of such cohorts solely for research purposes is the huge initial cost of setting up the infrastructure required to recruit a sufficiently large sample size. In this context, the recently initiated Department of Health NHS Vascular Health Check Programme provides a unique opportunity to establish such a cohort as all individuals in the appropriate age range (40-74 years) free of CVD are being invited to their general practices for a vascular health check. The large number of subjects that will be assessed in a systematic manner for cardiovascular risk and who will all have blood samples routinely collected for lipid profiling provides an ideal scenario to add a research project that can help determine whether a GRS will be useful in predicting CAD risk. ***In short, the Vascular Health Check programme provides a unique opportunity to test, in a major way and at a marginal cost, whether in the future adding genetic information will improve CAD risk prediction.***

Hypothesis

Addition of genetic information in the form of a genetic risk score based on recent discoveries of genetic variants that associate with CAD will improve risk prediction of coronary artery disease.

Objectives

The objective of the **GENVASC Study** is to recruit (with informed consent) subjects who attend their GP practices to have a vascular health check. Subjects will be asked to consent to:

- providing blood samples for research (in most cases taken at the same time as clinical samples for their vascular check)
- allow the GENVASC research database to hold semi-anonymised (no identifiable names) data on their CVD risk score calculated using conventional algorithms
- to allow the database to be periodically updated with any CVD outcomes via GP practice databases as well as from appropriate national registries
- to allow their stored samples be used for DNA and other analyses (again anonymised).

The GENVASC Study will not interfere with the primary clinical imperative of the Vascular Health Check Programme. It will **retrospectively** analyse whether addition of a GRS would have improved CAD risk prediction. The number of subjects required and the power of the study are described in Appendix 1.

Feasibility

This is a highly opportune time to undertake the GENVASC Study in Leicester for several reasons:

- The Leicester City Commissioning Group have prioritised the delivery of vascular health checks as a high priority and plans to invite over 80,000 people for screening in the next 3 years. They have strongly endorsed the development of the GENVASC Study as an important addition to their plans and will fully support its implementation.

- The GENVASC Study will be a core translational research project within the NIHR Leicester Cardiovascular Biomedical Research Unit which has recently received renewed funding for a further 5 years for £6.5M from the Department of Health. The GENVASC project was included as one of the new projects that will be conducted within the "Cardiovascular Genetics" theme of the BRU. However, please note that we will be seeking portfolio adoption and CLRN support to meet the non-BRU costs incurred in GP practices.
- The new £12.5M University of Leicester Cardiovascular Research Centre at Glenfield supported by the British Heart Foundation is developing state-of-the-art robotised facilities for large-scale (> 250,000 samples) sample processing and storage (The Leicester Cardiovascular Bioresource) and will also house the outstanding informatics infrastructure (BRICCS) that we have developed under the BRU to collect, harmonise, store and analyse data from multiple sources.

This coalescence of opportunities and strengths makes GENVASC a highly feasible project for us to deliver, despite its ambitious scale. The GENVASC study provides an excellent opportunity for wider engagement of the clinical community in Leicester in research. Indeed, if we can show its feasibility and cost-effectiveness, then we anticipate rolling it out into practices covered by other CCGs locally as well as discuss where the NIHR wishes to adopt it more broadly. ***The GENVASC Study has the potential to be a flagship project for the NIHR.***

GENVASC Study protocol

A diagram showing the flow of the GENVASC Study is shown in **Figure 1**.

Project Board: The GENVASC Study will be managed by a project board comprising the Principal Investigator and co-investigators and will include a patient representative. Other members will be co-opted as required. A specific **project manager** will be appointed who will be responsible for its smooth running, including bringing practices on board (see below), providing training to nurses (see below) and liaising with staff in the BRU.

Subjects: Inclusion and exclusion criteria will be minimally restrictive. Inclusion criteria include any adult subjects between the ages of 40-74, who consents to participate in the project. The main exclusion criteria will a known history of CVD or a known history of blood transmissible infection (Hep B, HIV). The initial target population is subjects attending their general practices for a vascular health check in Leicester PCT. The ambition set by Leicester PCT is to try and screen all subjects (n > 80,000) between 40-74 years of age without a known history of CVD in the next five years through a process combining both opportunistic screening (if the patient attends the practice for another reason) or through invitation to attend for a check. In future, recruitment may also be extended to practices in other local PCTs or other healthcare settings using the same protocol.

Duration of recruitment and sample size: The GENVASC Study initially proposed to recruit 15,000 subjects (see power calculations in Appendix 1). The recruitment has gone extremely well and we have to date recruited just over 20,000 subjects. Because of the potential to recruit even more subjects across the East Midlands and having secured the new BRC funding we now propose to continue to actively recruit until March 2021 with a target sample size of in excess of 30,000 individuals. The rationale for this is that the robustness of any conclusions we draw are entirely dependent on accrual of cardiovascular events in the cohort and the larger sample size will allow more outcomes to occur and be collected within a timely fashion.

Recruitment and Patient Information: Patients attending the NHS Health Check will be provided with study information either at or in advance of their appointment.

Where the witnessed consent process is applied the participant will be provided with an abbreviated information sheet and the study will be explained verbally with a supporting full information sheet and

withdrawal form. Where the unwitnessed consent process is applied the participant will be provided with the full information sheet, consent form and withdrawal form combined with verbal confirmation of consent and discussion of the study.

Generally consent is provided prior to sampling of clinical bloods such that no additional venepuncture is required but in some cases, usually where it is more convenient to the participant, an additional venepuncture is required to collect research bloods subsequent to the clinical bloods.

The practice staff will receive full briefing on the project from the Project Team and training to answer any questions and to obtain informed consent. The PIS will be available in multiple languages.

Consent: If a patient is happy to participate, written consent will be obtained. A copy of the consent form will be provided to the patient and a copy also stored in their primary care records. A third copy will be kept centrally by the investigators. Subjects will be asked to provide consent:

- to take 20 mls of blood for research
- for the LCB research database to hold pseudo-anonymised (no identifiable names) data on their CVD risk factors and risk score determined at the time of their vascular health check and other relevant information
- for the database to be periodically updated with any CVD or other outcomes from records held in their GP practice databases as well as from appropriate local (e.g HES data) and national registries (e.g MINAP)
- for their stored samples be used for DNA and other analyses (again pseudo-anonymised)
- for them to be contacted for future studies (optional)

Withdrawal: Patients will also be provided with details on how they can withdraw from the Study at any stage if they wish and that they have no obligation to participate in any future studies they are approached about. Both the consent form and the PIS will be printed in multiple languages.

Samples: 20 mls of blood will be obtained from each participant for research. This will include an EDTA sample for DNA extraction and plasma as well as a serum sample. Most samples will initially be sent to the Pathology Dept at UHL (together with any clinical samples). A system will be established to transfer the research samples to the BRU from UHL pathology. In some circumstances study samples will be shipped directly to the BRU. The samples will be processed, aliquoted and stored in the LCB facility at Glenfield Hospital using a unique Study identifier for each patient (see below regarding anonymisation).

Data and Anonymisation: A GENVASC database will be created as part of the BRICCS Informatics platform of the BRU. The database will contain data (see below) on each subject pseudo-anonymised on a **unique Leicester Cardiovascular Bioresource identifier number**. Under this ID, the database will contain information on each subject's demographics, their cardiovascular risk factors as well as their CVD risk score, and any other relevant information extracted from the GP practice database. With time, the database will also be populated by additional information including data on outcomes as well as results of additional investigations such as further cholesterol levels that the patient has, as well as findings from research analyses of the samples. Methods for validation of key outcomes through either examination of clinical notes or cross-referencing of databases will be established. Access to the GENVASC databases will be restricted to bona fide investigators (all of whom will have had GCP training) by password restriction and real-time monitoring of who accesses the data. Ability to add or modify the data will be restricted to key personnel.

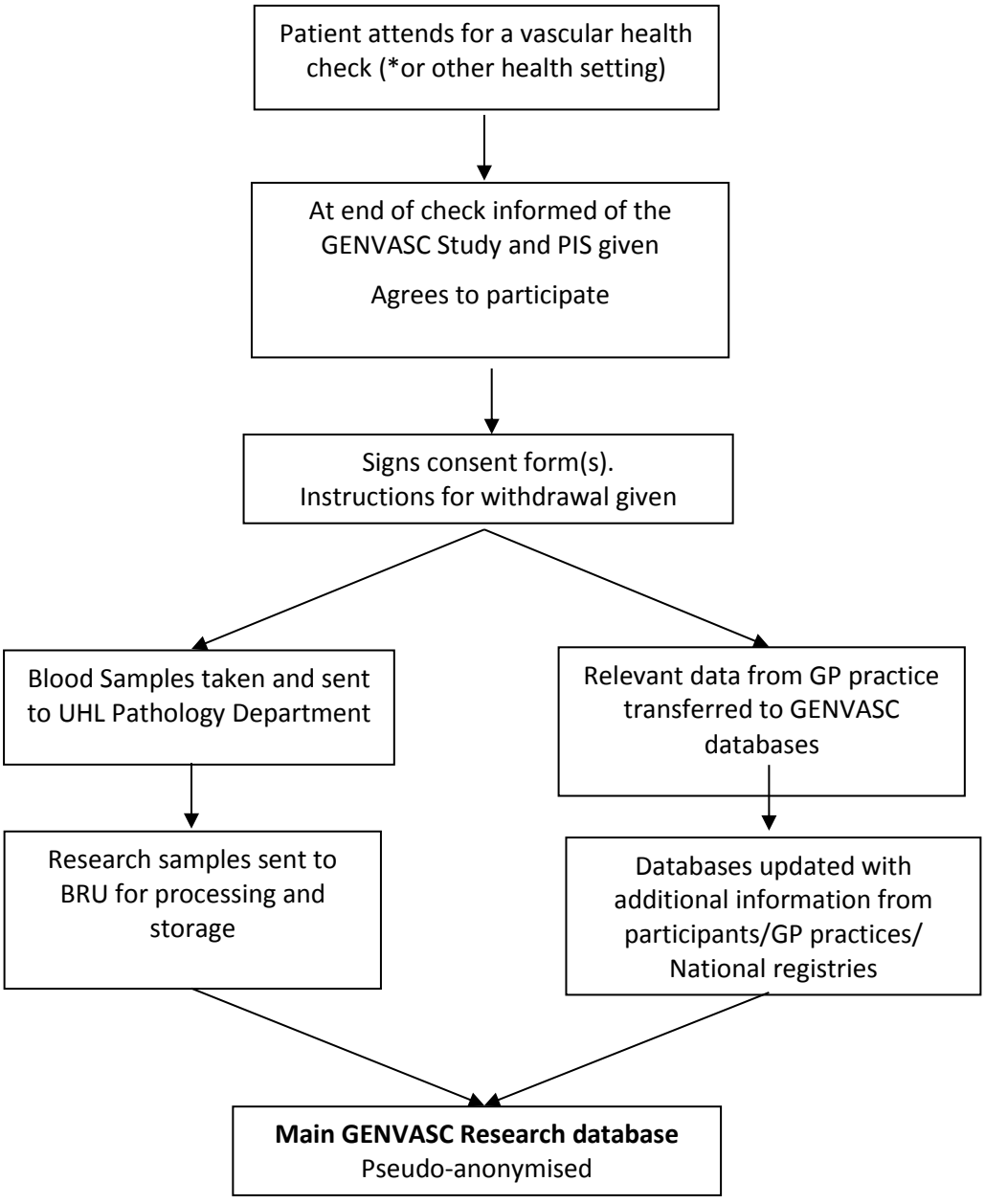
To allow updating of the GENVASC database with clinical data to occur while maintaining pseudo-anonymisation, a separate database will be maintained containing the LCB identifier and the patient's unique NHS S Number. Data will be extracted from various clinical sources using the NHS S Number into this database and transferred to the main GNVASC database using the matching LCB identifier. Once this transfer has happened, the extracted data will be destroyed. If a patient wishes to withdraw from the study and asks for their sample and data to be destroyed, this can also be achieved easily using this method.

Laboratory analysis: Samples will only be identifiable by their LCB ID number. EDTA samples will be used to extract DNA and plasma. Genetic analysis will be performed locally or through collaboration with national

or international collaborators or through commercial genotyping companies as dictated by the needs of the study and what is most cost-effective. Data from such analysis as well as from analysis of the plasma/serum samples for biomarkers will be stored in the GENVASC database. At no stage will those undertaking such analysis have information that could enable them to identify individual participants.

Access to data and samples: A LCB steering committee will be established to approve cardiovascular projects that wish to access data and samples beyond those needed for the primary GENVASC Study objective on the incremental benefit of genetic information in predicting CAD. All projects will also be reviewed by the Patient and Public Involvement (PPI) Committee of the BRU. Access will be restricted to data and samples relevant for the project and in a manner that cannot identify individual patients (unless appropriate). A condition of access will be that researchers deposit any findings they generate on samples provided to them in the LCB databases once the findings have been published (or at an agreed time), so that the data can be available to other researchers.

Figure 1: Flow diagram of the GENVASC Study



*GENVASC invitation, PIS and unwitnessed consent form included in the GP practice invitation letter for NHS health Check

Appendix 1: Sample size considerations and statistical analysis (Prepared by Professor John Thompson)

The purpose of CAD risk prediction is to classify individuals into different risk categories so that preventative efforts, especially expensive ones, can be targeted at those likely to benefit most. The cost-effectiveness of risk prediction depends on the accuracy of this risk categorisation and for a common condition such as CAD even a relatively small improvement in the accuracy of risk prediction can have significant public health implications and benefit. The purpose of GENVASC is to test whether the addition of genetic information in the form of a genetic risk score (GRS) can significantly improve CAD risk prediction over current methods.

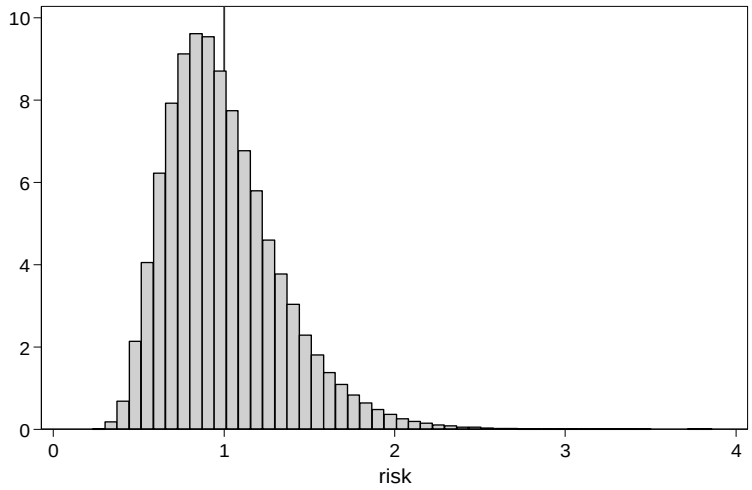
A widely used risk score for CAD is the Framingham risk score, in which individuals are classified as at either low (< 10%), intermediate (10-20%) or high (> 20%) risk of CAD at 10 years based on a number of characteristics including age, gender, smoking status, history of hypertension or diabetes and serum cholesterol level. In NHANES III survey 73% of individuals between 40-79 years were classified at low, 23% at intermediate and 4% at high risk, respectively.

Based on these figures we simulated a sample of 15,000 individuals (the size of cohort we hope to recruit in GENVASC):

Framingham group	Freq.	Percent	Cum.
Low	10,893	72.7	72.7
Intermediate	3,489	23.2	95.9
High	618	4.1	100.0
Total	15,000	100.00	

We then assigned genotypes at 48 SNPs (the current number of independent validated CAD variants) using allele frequencies and odds ratios determined by the CARDIoGRAM+C4D consortium in over 190,000 individuals. This had the effect of increasing the risk for some and decreasing it for others (**Figure 1**)

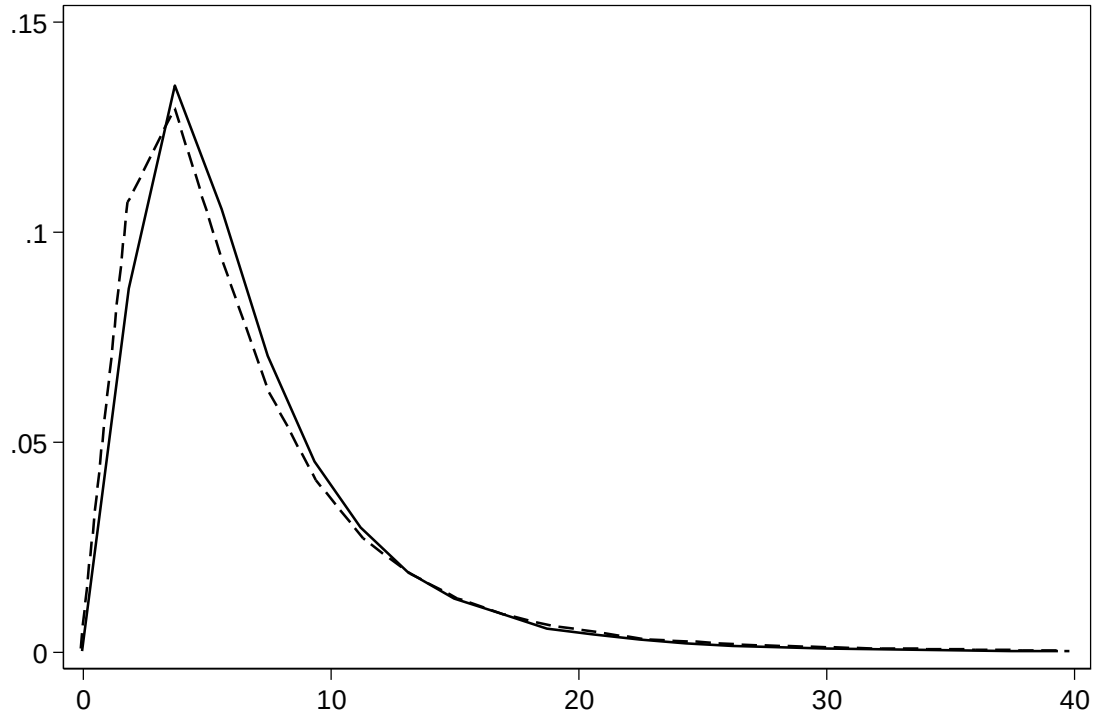
Figure 1



Half the population have a genetic risk between 0.76 and 1.18 (hardly changed); but 2.4% have their risk estimate halved (or more) and 1.1% have it doubled (or more)

If we assume that the genetic risk and the Framingham risk are independent then combining them creates a new risk score. The distribution of CAD risk scores before (solid) and after (dash) are shown in **Figure 2**.

Figure 2



The distributions of two scores differ very little although after including the genetic information the distribution is slightly more spread out. This is precisely what has been observed in practice (e.g. see Schunkert H et al. Nat Genet 2011).

The numbers in each category now are

Framingham+GRS	Freq.	Percent	Cum.
Low	11,228	74.9	74.9
Intermediate	2,967	19.8	94.6
High	805	5.4	100.0
Total	15,000	100.00	

Overall, the proportions in each category are not much changed (c Table above) although the movement is slightly towards the extreme categories.

Comparing the movement of *individuals* between categories before and after the addition of the GRS:

	Framingham + Genetic Risk Score			
Framingham	Low	Intermedi	High	Total
Low	10,227	659	7	10,893
	93.88	6.05	0.06	100.00
	91.09	22.22	0.87	72.62
Intermediate	996	2,122	371	3,489
	28.55	60.81	10.64	100.00
	8.87	71.51	46.12	23.26
High	5	186	427	618
	0.75	30.16	69.10	100.00
	0.04	6.28	53.01	4.12
Total	11,228	2,967	805	15,000
	74.85	19.78	5.37	100.00
	100.00	100.00	100.00	100.00

So of those originally in the high risk category ~ 69% (427/618) stay at high risk, 31% move to a lower risk category. Of those who end up in the high risk category, ~ 53% (427/805) were already high risk, ~46% (371/805) were previously intermediate risk and ~1% were previously low risk. Of subjects initially in the low risk group, ~16.5 % (659/10,893) move into intermediate risk

So although the risk score over the population changes very little (as we have seen in previous risk score analyses) at the individual level the impact can be more substantial.

In the simulated group of 371 people classified as intermediate by Framingham but High by Framingham + Genetic Risk Score, the Framingham risk averages 15.9% over 10 years while adding the genetic information increases this to 24.3%. The power to reject 15.9% in a two-sided test at the 5% level would be **98% after 10 years** and even at 5 years the power would still be 80%. If we were able to only recruit 10,000 subjects, the power would still be acceptable (**65%** at 5 years and **92**% at 10 years).

Conversely there are 186 people classified as high risk by Framingham but Intermediate risk when the genetic score is included. The average risk for these people under Framingham is 23.1% over 10 years while their true risk (under Framingham + Genetic Risk Score) is 16.1%. The power to reject 23.1% at 10 years is **76%**.

Power estimates based on movement of individuals **between** absolute risk categories, while perhaps more relevant clinically, underestimates the full potential of the data. In 10,000 people we would expect there to be 526 people whose risk assessment increases by more than 5% when we take GRS into account. The simulation suggested that over **5 years** the Framingham score would give them an average risk of 7.5% while the GRS increases that to an average 12%. The study would have 97% power to confirm this change with the events occurring over 5 years. In the opposite direction 361 would be expected to have their risk assessment lowered by more than 5% - the Framingham risk for these people would average 9.7% and we would have 82% power to reject that value in a one-sided test

These calculations are based on a number of assumptions: (i) that the genetic risk score is independent of the Framingham score (ii) that the effects attributed to each variant are accurate (iii) that there isn't a substantial loss of CHD events due to mortality from non-CHD causes (iv) that the proportion of subjects in the different Framingham risk categories in GENVASC are similar to those in NHANES III. Each of this has a different impact on power. About 35% of the CAD SNPs identified to date are also associated with a traditional CAD risk factor (mainly lipids or BP) and the impact of these could have been accounted for in the Framingham Score. This would reduce power. On the other hand, genetic variants are present from birth and their impact on life-time risk (even if they work through a traditional risk factor) may be significantly underestimated from cross-sectional analyses. Furthermore, it is likely, given the demographic and ethnic mix of Leicester, that there will be a greater proportion of individuals at intermediate or high risk compared with NHANES III. While the exact impact of these factors is difficult to estimate, they would tend to increase power.

Taking these issues into consideration, we believe that a study of **10000** subjects and certainly one of **15,000** subjects would provide robust data (especially with time and more events) as to the benefits of adding a GRS to current algorithms for risk prediction. The power may be even greater as more genetic variants, especially rare variants with stronger effects, emerge from on-going discovery efforts.

Statistical analysis

The data from this study will be summarised in terms of the number of first CAD events and the number of person-years of follow up. These will be modelled using Poisson regression in order to allow for the base line characteristics of the patients at recruitment. This will enable us to compare the observed event rate with the expected number as predicted by the Framingham Risk score and also with the expected number as predicted by the Framingham score plus genetic risk score (GRS). We will test for improved prediction over the entire sample when the GRS is added and will also look at the benefit from GRS use in the higher risk patients who would be targeted for treatment. Secondary analyses will explore the association between the risk variants and other known risk factors such as blood pressure and cholesterol and with specific outcomes such as MI and mortality. As the database grows it will also be possible to explore the changing pattern of CAD risk with treatment or with alternations in risk factors using event history models.