Imperial College London

Full title: Investigation of lifestyle, environmental, genomic and molecular factors underlying health outcomes in South Asians and Europeans – the LOLIPOP 100K study

Short title: The London Life Science Population (LOLIPOP) 100K Study

Protocol – version 2
11 January 2021

IRAS reference: 248261

Sponsor:
Imperial College London is the main research Sponsor for this study. For further information regarding the sponsorship conditions, please contact the Head of Regulatory Compliance at:

Joint Research Compliance Office
Imperial College London and Imperial College Healthcare NHS Trust
Room 215, Level 2, Medical School Building Norfolk Place London, W2 1PG

Tel: 0207 594 9459/ 0207 594 1862
http://www.imperial.ac.uk/joint-research-compliance-office

Key contacts

Chief Investigator
Professor Jaspal S Kooner
Professor of Clinical Cardiology, Imperial College London
email: j.kooner@ic.ac.uk

Co-investigator
Professor John Chambers
Professor of Cardiovascular Epidemiology, Imperial College London
email: john.chambers@ic.ac.uk

Programme Management and Study co-ordination
Mr Christopher Brookes
National Heart and Lung Institute
Imperial College London
email: c.brookes@ic.ac.uk

Imperial College London

SIGNATURE PAGE

The undersigned confirm that the following protocol has been agreed and accepted and that the Chief Investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the principles outlined in the Declaration of Helsinki, the Sponsor's SOPs, and other regulatory requirement.

I agree to ensure that the confidential information contained in this document will not be used for any other purpose other than the evaluation or conduct of the investigation without the prior written consent of the Sponsor.

I also confirm that I will make the findings of the study publicly available through publication or other dissemination tools without any unnecessary delay and that an honest accurate and transparent account of the study will be given; and that any discrepancies from the study as planned in this protocol will be explained.

For and on behalf of the Study Sponsor:

Signature: Date:

Becky Ward Digitally signed by Becky Ward Date: 2021.01.19 18:13:43 Z

XXX

.....

Becky Ward
Research Governance Manager Imperial College London

Chief Investigator:

Signature: Date:

XXX

.....

Professor Jaspal S Kooner
Imperial College London

Funding and other support

Support for the proposed programme of work has been obtained through

- Wellcome Trust (£5M, 2018-2023; ref: 212945/Z/18/Z)
- National Institute for Health Research (NIHR, £7M, 2018-22, ref: 16/136/6).
- The North-West West London NIHR Clinical Research Network who have agreed to provide organisational support, including facilitating letters of invitation to people registered with local GPs.
- General Practitioners in West London, including the Clinical Care Groups of Ealing, Hounslow, Brent and Harrow, and Hillingdon

1. Study summary

Diabetes and cardiovascular disease are 2-4 fold more common in UK South Asians than Europeans. The reasons underlying this health disparity are not known. To advance understanding of metabolic, cardiovascular and other important chronic diseases in UK South Asians, we propose to establish LOLIPOP 100K, a longitudinal population study comprising ~100,000 South Asian and European men and women, aged 25-85 years, under long-term follow-up for health outcomes. We will deliver this by building on our previously completed initial LOLIPOP study, comprising 30,000 people and carried out in 2002-8. In this new phase we extend this previous work and recruit up to 100,000 new South Asian men and women to create LOLIPOP 100K. Recruitment will be via GP lists and coordinated centrally to allow targeting of South Asians and Europeans, and enable monitoring of participation by factors including age, sex, socioeconomic status. All participants will be asked for their consent to participate and will have baseline assessment moving through a series of 'assessment stations' including questionnaires, physical measurements and blood/urine sampling. Baseline assessment will be of ~ 90 mins duration. Assessment will be carried out by staff (nurses / health care assistants) trained specifically for the LOLIPOP 100K Study. Data will be collected using a customised, integrated IT system. Blood samples will be processed in the central laboratory, and aliquots will be stored at -80°C for future analysis.

Participants will be followed-up for their health long-term through medical and other health-related records. We aim to use a range of data sources including linkage for: i. death; ii. site-specific cancer, iii. Hospital Episode Statistics (HES) for England and Wales; Scottish Morbidity Record (SMR) for Scotland; iv. myocardial infarction (MINAP), stroke (SINAP); v. pathology services at local hospitals; vi. hospital databases (CVD, T2D, ESRD, histopathology); vii. primary-care electronic health records. The data and samples will be used to investigate the behavioural, environmental, genetic and other molecular factors influencing health and disease in South Asian populations.

2. Background

Importance of cardiovascular disease and diabetes in South Asian populations

Cardiovascular disease (CVD) and type-2 diabetes (T2D) represent major global health challenges.(1-3) CVD accounts for 17.9 million lives each year representing ~ 31% of all global deaths. An estimated 500M people have T2D world-wide. South Asians who comprise ~1/4 of the world's population have a disproportionally high burden of CVD and T2D.(4-6) In the UK, people from South Asian ethnic groups make up ~5% of the population. National mortality statistics show that CVD mortality is ~2-fold higher among UK South Asians than Europeans, observed in all major South Asian subgroups.(7) Notably, CVD death rates are ~4 times higher in younger South Asians, compared with Europeans. CVD mortality has declined in UK for both men and women by ~2/3 over the last 30 years.(4) However, the rate of decline has been slower among South Asians. UK South Asians have an ~3 times elevated risk of T2D compared with Europeans,(8) and typically develop the disease 5-10 years earlier. The high burden of CVD and T2D amongst South Asians represents an important health inequality, which needs to be addressed urgently.

Mechanisms underlying increased CVD in South Asians

Traditional risk factors including cigarette smoking, hypertension and hypercholesterolaemia contribute to CVD in all populations, including South Asians.(9) However, numerous studies show that conventional CVD risk factors are not more prevalent among South Asians, and do not explain their increased CVD risk, compared with Europeans.(4, 10) In contrast, T2D, insulin resistance and related metabolic disturbances (high triglycerides, low high-density lipoprotein cholesterol, central obesity) are more common amongst South Asians.(4, 10) The Southall and Brent Revisited (SABRE) Study,

a prospective comparison of CVD mortality among UK South Asian and European men (*n* = 1,420 and 1,787 respectively), reported ~2-fold excess CVD risk amongst South Asian men, which persisted after adjustment for obesity, diabetes, insulin resistance, blood lipids, blood pressure and smoking.(11) Though SABRE provided evidence that ~2-fold higher incident CVD in South Asians was not accounted by known risk factors it had many limitations. Sabre did not; i. include women; ii. investigate the role of many key exposures including diet and physical activity, and; iii. did not examine the role of emerging or genomic risk factors. Reasons underlying the high CVD risk amongst South Asians thus remain poorly understood, and this knowledge gap represents a major obstacle in risk stratification and prevention of CVD in this population.

Improving risk stratification for CVD among South Asians

Effective prevention of CVD requires robust methods for the identification of people at increased risk. Framingham and other prospective studies provide multivariate equations to predict risk of CVD.(12) However, these functions remain unvalidated in South Asian population. Cross-sectional data indicate that Framingham risk functions underestimate CVD risk in South Asians.(13, 14) One possible explanation is that the Framingham risk functions do not adequately incorporate insulin resistance, and related metabolic disturbances, which are more prevalent in South Asians. The failing of the risk functions to predict CVD mortality accurately in ethnic minority groups has been repeatedly recognised as a major limitation, with consequences for risk prevention among South Asians. Underestimation of risk will lead to under-treatment, and widen the inequalities in CVD outcomes for this vulnerable population. The National Institute for Clinical Excellence, and the Joint British Societies (British Cardiac Society, British Hypertension Society, British Diabetic Association) have all called for research to develop and validate new risk models that accurately predict CVD risk amongst South Asians.

3. Rationale for the current proposal

The London Life Science Population (LOLIPOP) study

To advance understanding if the mechanisms underlying CVD and T2D in South Asians,, we established the London Life Science Population study in (LOLIPOP, 2002-8). This initial, phase 1 study comprised a longitudinal population study comprising 32,183 South Asian and European participants, aged 35-75 years recruited from the lists of 58 general practitioners in West London. The study was approved by NRES committee – London Fulham (07/H0712/150), and has been supported through competitive grants awarded by British Heart Foundation, National Institute for Health Research, Wellcome Trust and the Medical Research Council. This initial LOLIPOP study comprised a focused self-administered health and lifestyle questionnaires, physical measurements relevant to cardiometabolic disease, and storage of a limited repertoire of biological samples (whole blood, serum, plasma and urine) at -80^{o}C for molecular analyses. Record-linkage was been established to routine NHS data, and has enabled identification of 1,297 deaths, as well as new cases of T2D (n=2,015), CVD (n=1,832), cancer (n=885) and end-stage renal disease (ESRD, n=58). The LOLIPOP study dataset has proved a rich and valuable resource for investigation of the genetic and environmental mechanisms underlying CVD, T2D, and related metabolic disturbances in South Asians. The cohort study has contributed to identification of hundreds of genetic variants influencing risk of CVD, T2D, insulin resistance, obesity, and lipid levels among South Asians.(15-22) Our recent observations include findings of defects in gene regulation in obesity that precede the development of T2D.(23). Our results have had translational impact, including identification of waist circumference and HbA1c as simple, readily accessible, clinical biomarkers that accurately predict future T2D in South Asians. These observations have underpinned international clinical trials aimed at reducing the risk of diabetes in this high-risk population.

The need to create the LOLIPOP 100K study

Disease risk results from a complex interplay of genetic susceptibility lifestyle and environment factors. Though LOLIPOP is well powered to identify the factors influencing common diseases such as CVD and T2D in South Asians, there remain important limitations. The current sample size of ~30,000 South Asians and Europeans is underpowered for investigation of CVD sub-phenotypes such as stroke, myocardial infarction, heart failure and renal failure, that are more common in South Asians than Europeans. Despite its size, LOLIPOP will yield insufficient numbers of disease cases in the younger age groups (age 25-35 years), and among women, which may yield unstable estimates (eg. due to random variations). LOLIPOP baseline phenotyping did not include formal assessment of dietary habit, physical activity, lung function, visual health or cognition. Incomplete or inadequate measures of potential risk factors and confounding factors are known to yield under- or over-estimates of disease. Recognising these limitations, there is an important need to create LOLIPOP 100K, comprising increased sample size and a wider range of exposures and biological samples. Longitudinal and detailed follow-up of participants with routine medical and other health-related records will allow identification and validation of a wide range of conditions necessary to reliably quantify the role of various risk factors underlying major diseases in this population. In recognition of this proposal, The Wellcome Trust and the National Institute of Health Research (NIHR) have now awarded programme grants (total ~£12M), to create the LOLIPOP 100K study, in order to advance understanding of the aetiological factors underlying the major health outcomes relevant to UK South Asians. The new LOLIPOP 100K resource will provide a unique repository of data and samples for chronic disease research among South Asians.

4. Research aims

The primary aim/objective of our research is to identify the genetic, environmental and lifestyle factors influencing health and disease, including diabetes and cardiovascular disease in South Asian populations, and to use the results to improve prediction and prevention of chronic disease.

There are no secondary objectives for this study

5. Entry Criteria

Inclusion: Men and women age 25-85 years (oversampled for South Asians)

Exclusion: Unable to give informed consent
Known to be pregnant
Current or recent acute illness

6. Study Methods

Design

To address our research aims, we propose to create the LOLIPOP 100K study, an observational, longitudinal population study comprising up to 100,000 South Asian and European men and women, aged 25-85. Participants will be undergo comprehensive phenotyping, including collection of questionnaire, physical measurements, retinal imaging, and biospecimen collection. Participants will continue to be followed for at least 20 years after the last subject has been recruited to identify future health outcomes. The LOLIPOP 100K resource created will provide a powerful epidemiological tool for investigation of the risk factors and mechanisms underlying health and disease in UK South Asians.

Recruitment

Our approaches to participant recruitment will be based on our successful experience in recruiting ~30,000 participants to the completed LOLIPOP study. Recruitment will be carried out in partnership with GPs in West London and other UK regions with high South Asian populations.

Potential participants will be South Asian (originating from India, Pakistan, Sri Lanka, Bangladesh) and European men and women aged 25-85 years. There will be oversampling of South Asians to achieve an expected sample size of up to 80,000 in this ethnic group. We anticipate recruiting up to 20,000 European participants. When recruiting through postal mechanisms we anticipate that there will be some participants of other ethnic groups who are enthusiastic to take part; these individuals can provide useful information on risk factor profiles in Non-Asian ethnic groups and so will be permitted to take part.

Potential participants will identified from GP practice lists and invited to attend an appointment at dedicated LOLIPOP 100K research clinics. SMS text invitations will be sent out to potential participants by the respective GP (in partnership with the Primary Care Research Network) and depending on response rate letters of invitation will also be sent to potential participants by the respective GP (in partnership with the Primary Care Research Network). The invitation letter will include options of date and time of appointment. Potential participants will be asked to confirm their appointment by; i. telephoning the central booking office; ii. email; iii. mailing the reply form in the pre-paid envelope provided or; iv. on-line visiting the study website. Potential participants who agree to attend for assessment will be sent a written confirmation of their appointment, along with advice on preparations for attending the assessment centre, and to record information that they might have difficulty recalling during the visit (e.g. medications, operations). We anticipate a response rate of ~50%, based on previous experience in the completed LOLIPOP study, and thus to invite ~ 200,000 people to take part, to achieve our goal of recruiting 100,000 people to the LOLIOP 100K study.

The central booking office will be available to confirm, change, cancel appointments and allow questions from potential participants to be addressed (e.g. transport and parking; travel expenses; assessment centre procedures; consent and withdrawal; feedback of results; confidentiality). Potential participants who confirm an assessment visit appointment but then do not attend will be sent a letter within 1-2 weeks of the missed appointment. This will ask them to contact the freephone service to book another appointment if they might still like to participate.

Knowledge about the research will also be promoted through community activities including open meetings with the public, meetings with community organisations and our study website. Members of the community who are interested to participate in the research, and who meet study entry criteria, will have the opportunity to register for the study.

Study visit

Assessment centres will be in hospital or community locations with good public transport links, proximity to parking, and disabled access. Usual opening hours will be typically Monday to Friday 8.00 am to 8.00 pm and Weekends 8.00 am to 5.00 pm. Assessment centres will be staffed by nurses, healthcare assistants, technicians and receptionist and senior nurse manager. All staff will be fully trained in baseline assessments. Participants will be asked to attend for a health evaluation after an overnight fast (water only after midnight). All participants will complete a structured assessment by the research team, comprising: i. Registration and consent; ii. Questionnaire; iii. Physical measurements; iv. Biological samples and v. Physical activity monitoring. The baseline assessment expected to take approximately two hours.

i. Registration and consent

Written, informed consent will be obtained from all participants for data collection, and inclusion in the research. Given the approach to recruitment, all participants will have had at least 24 hours to consider whether to take part. Informed consent will include permission for the data and samples collected to be used for chronic disease research, including data sharing with national and international bodies concerned with prevention and control of T2D and CVD, as well as for molecular epidemiological research. Participants will be given the opportunity to ask questions during the consent discussion. Multilingual research facilitators fluent in the major South Asian languages (Punjabi, Hindi, Urdu) will be available as necessary to facilitate collection of informed consent. Participants unable to give informed consent will not be recruited.

ii. Questionnaires will comprise:

a. Health and lifestyle questionnaire will be used to collect information on behavioural risk factors (smoking, alcohol habit, physical activity), personal and family medical history, medications, socio-economic status and other measures of lifestyle and environmental exposure.
b. Dietary habit by interviewer administered 24-hour dietary recall, using the online digital platform "INTAKE24" (https://intake24.co.uk/). This does not store identifiable data.
c. Cognition will be measured using a validated, digitally delivered tool, originally developed for UK Biobank. The cognitive assessment asks a series of question that assess memory and thinking, as measures of cognitive performance. The assessment takes ~15 minutes. This does not store identifiable data.

i. Physical measurements will comprise:

a. Height (SECA213 Stadiometer or equivalent)
b. Weight (OMRON BF-511 or equivalent)
c. Waist and hip circumference (non-stretchable SECA201 measuring tape or equivalent)
d. Bio-impedance for body fat composition (OMRON BF-511 or equivalent)
e. Blood pressure by digital device (OMRON HEM-9210T or equivalent)
f. Resting 12 lead ECG (MAC 2000 ECG machine or equivalent)
g. Lung function will be measured by spirometry (NuvoAir spirometer or equivalent). Each participant will be asked to complete three measurements.
h. Retinal imaging. A photograph of both retina will be taken using a non-mydriatic retinal camera (CrystalVue700 or equivalent). This device captures images of both eyes in an automated fashion, in under 2 minutes, and *without pupil dilatation*.

ii. Biological samples will comprise:

a. Venous blood (total 60mls, by standard aseptic technique, from the antecubital fossa)
b. Urine sample (spot urine)
c. Stool sample (optional, 10-20g collected from fresh stool sample).

iii. Assessment of physical activity

Physical activity will be measured by wrist-worn accelerometer device for 7 days (AX3 activity tracker), as used in the UK Biobank study (https://www.ukbiobank.ac.uk/activity-monitor-3/). The device is returned by pre-paid post to the research centre.

A fully integrated IT system will allow direct electronic data capture. Subject to Information Governance team review and approval, we propose a cloud-based research database, using a platform that is GDPR and NHS toolkit compliant. All data will be captured electronically into computers via keyboards, touch-screens, bar-code readers and direct transfer from measurement devices (eg, the electronic sphygmomanometer or spirometer).

7. Participant risk and burden

The assessments proposed are all non-invasive, well established and safe. They have all been successfully used in large scale population studies such as UK Biobank, involving >500,000 Europeans, and have been shown to be feasible and acceptable to participants. The research team have considerable experience in the assessments proposed through their other completed and ongoing population research studies.

8. Sample management

Biological samples (blood, urine and stool) will be collected from participants. Aliquots of whole blood, serum, plasma and urine will be stored for future quantification of biomarkers for cardiovascular and metabolic health (eg measurement of lipid profile, renal function and insulin levels), and as a resource for future molecular epidemiological studies to investigate the mechanisms underpinning the development of CVD, T2D, and other complex disease that are of importance to South Asian populations (including but not limited to: obesity, cancer, dementia, COPD, chronic kidney disease). The molecular epidemiological research will employ a range of strategies including, but not limited to: genome sequencing and genotyping, studies of genomic regulation (eg quantification of DNA methylation), molecular profiling of serum and urine (eg by targeted and untargeted metabolomics, or proteomics).

The aliquots of blood, urine and stool will be stored long term (>20 years) at -80C or below. All storage locations will be physically secure (eg 24 hour security staff and CC-TV), with continuous monitoring of and recording of freezer temperatures, air-conditioning and remote alarms in the event of freezer failure. To help ensure the security of this unique resource, sample aliquots from each participant will be split and stored in separate locations.

9. Incidental Clinical Findings

LOLIPOP 100K participants will be offered a report comprising physical measurements (adiposity, blood pressure), biochemistry (lipid profile, glucose, HbA1c) and clinically reported 12 lead ECG. The rationale for our approach is that these tests that are recommended by the NHS as part of routine, periodic health-check in the adult population, and are an important part of our commitment to stakeholder involvement. Our experience from the LOLIPOP study is that these health reports were well received by the participant and their GP, and were considered a valued benefit from taking part in research.

We do not intend to routinely report the new phenotypes being measured (ie lung function, retinal imaging, cognition or diet). However, if a clinical abnormality of major significance is noticed during the course of the study visit (eg retinal detachment on fundus image), this will be reviewed by the study team for consideration of reporting to the participant as an incidental clinical finding. A similar approach has been adopted by UK Biobank for their imaging studies.

10. Adverse events

DEFINITIONS

Adverse Event (AE): any untoward medical occurrence in a patient or clinical study subject.

Serious Adverse Event (SAE): any untoward and unexpected medical occurrence or effect that:

- **Results in death**
- **Is life-threatening** – *refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe*

- **Requires hospitalisation, or prolongation of existing inpatients' hospitalisation**
- **Results in persistent or significant disability or incapacity**
- **Is a congenital anomaly or birth defect**

Medical judgement will be exercised in deciding whether an AE is serious in other situations. Important AEs that are not immediately life-threatening or do not result in death or hospitalisation but may jeopardise the subject or may require intervention to prevent one of the other outcomes listed in the definition above, should also be considered serious.

REPORTING PROCEDURES

All adverse events directly arising from attendance in the research will be reported. Depending on the nature of the event the reporting procedures below should be followed. Any questions concerning adverse event reporting should be directed to the Chief Investigator in the first instance.

Non serious AEs
All such events, whether expected or not, should be recorded.

Serious AEs
An SAE form should be completed and faxed to the Chief Investigator within 24 hours.

All SAEs will be reported to the Fulham REC where in the opinion of the Chief Investigator, the event was:

- 'related', i.e. resulted from the administration of any of the research procedures; and
- 'unexpected', i.e. an event that is not listed in the protocol as an expected occurrence

Reports of related and unexpected SAEs should be submitted within 15 days of the Chief Investigator becoming aware of the event, using the NRES SAE form for non-IMP studies. The Chief Investigator must also notify the Sponsor of all SAEs.

Local investigators should report any SAEs as required by their Local Research Ethics Committee, Sponsor and/or Research & Development Office.

Contact details for reporting SAEs

jrco@imperial.ac.uk

CI email (and contact details below)

Fax: 020 8967 5740, attention Professor Jaspal S Kooner

Please send SAE forms to: Professor Jaspal S Kooner

Tel: 020 8967 5000 (Mon to Fri 09.00 – 17.00)

11. Follow-up and future recontact

Our prospective population study design proposes that participants be followed up to identify people who will maintain health or develop disease. We will then relate this to the extensive range of risk factors captured at baseline. This presents an invaluable opportunity to investigate the causes and natural history of a wide range of diseases and health conditions that affect South Asians. There are two broad approaches to follow-up that will be adopted:

• *Routine record linkage*. Participant consent will be sought to enable long-term follow-up for their health long term through medical and other health-related records. This will include linkage for; i. death; ii. site-specific cancer, iii. Hospital Episode Statistics (HES) for England and Wales; Scottish Morbidity Record (SMR) for Scotland; iv. myocardial infarction (MINAP), stroke (SINAP); v. pathology services at local hospitals; vi. hospital databases (CVD, T2D, ESRD, histopathology); vii. primary-care electronic health records. Since it is not possible to describe the electronic health records that will be available in the future, it is not possible to provide an exhaustive list of data sources that we will use. In general we will aim to access a wide range of systems (current and future) that hold health data as is usual practice in large scale population studies internationally.

• *Re-contact with participants*. Direct re-contact with participants provides a valuable means to capture reliable information on a wide range of risk factors and medical conditions. We will obtain consent from all participants for permission to re-contact them in future for a re-screening examination or for further evaluation based on clinical and laboratory characteristics (ie phenotype and/or genotype). This might involve attending a repeat baseline visit to the research health clinic. Follow-up using web based questionnaires may provide a cost-effective option for future contacts. Further IRB approval will be obtained before participants are recontacted for research purposes.

12. Statistical considerations

Data will be analysed using appropriate statistical software (eg STATA, SPSS). Continuous variables will be assessed for normality, and non-normally distributed variables will be transformed as appropriate or will be assessed using non-parametric tests if appropriate. The data and samples collected will enable both cross-sectional and longitudinal analyses, and thus to define the prevalence and incidence of clinical outcomes of interest respectively. We will quantify the relationship between the clinical outcomes of interest, and potential underlying exposures (lifestyle, environmental, genetic) risk factors using regression techniques. For the assessment of relationship between a continuous primary and secondary outcome we will carry out mixed models analysis of covariance (e.g., BP, HbA1c, blood cholesterol, serum creatinine) etc. Mixed models logistic regression will be used to assess the association between explanatory and categorical outcome variables (e.g., T2D, CVD, hypertension, hypercholesterolemia). Statistical significance will be inferred at $P<0.05$, using techniques (eg Bonferroni) to control for multiple comparisons where appropriate. Where complementary phenotypic or molecular information is available from the preciously completed LOLIPOP study, we will carry out joint analysis of the LOLIPOP and LOLIPOP 100K datasets.

The LOLIPOP 100K study provide high statistical power for identifying environmental and genetic factors as well as their interaction effect that contribute to incidence of diabetes and cardiovascular disease, and the outcomes for these diseases in UK South Asians compared to Europeans. Based on the disease incidence observed amongst the current South Asian participants of the LOLIPOP study, we estimate that 10 year follow-up of 100K South Asians will yield >10,000 deaths and cases of common diseases such as CHD and T2D, >5,000 cancer cases, including ~2,500 cases of breast cancer, and ~300 cases of uncommon

endpoints such as end stage renal failure. This affords >80% power to detect an odds ratio of as 1.1 (common diseases) to 1.5 (uncommon diseases) for exposures with prevalence 20%, at P<0.05 (Table). Data and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study, including the follow-up period.

13. Expected value of results

We will deliver robust data on the incidence and identification of underlying risk factors for CVD, T2D, cancer and ESRD among South Asians. The data generated from the resource will provide unique information on novel molecular biomarkers and biological pathways underlying these diseases, to help develop new predictive tools, therapeutic targets, and prevention strategies among South Asians. Societal impact will be ensured by inclusion of major groups of South Asians, in diverse settings, making our findings generalizable. Academic impacts will be achieved through; i. data sharing; ii. collaborations that advance understanding of disease pathways; iii. publications, iv. policy improvement; v. teaching, training. Economic impacts will be achieved by improved prevention and treatment of common diseases affecting large numbers of men and women of working age, and reduced healthcare costs.

14. Confidentiality, data protection and data management

We will commission a purpose built database to facilitate i. robust collection; ii. secure storage and iii. appropriate sharing of population health data,. The database will be commissioned under the supervision of the Chief Investigator and study Sponsor.

Data will be captured through tablets and laptop/PC devices in both offline and online (connected) modes. A range of measures, in-line with international best practice, will be taken to ensure the protection of data collected as part of the study. GDPR and Data Protection legislation will be taken into consideration. Appropriate authentication and access control mechanisms will be implemented. Data will be stored securely (both locally and in cloud based infrastructure). All data will be encrypted during transmission (eg via use of https), and stored using approaches that are compliant to globally recognised information protection standards. Track record will also be assessed to ensure that physical storage of the data is secure. This will be described in a formal Data Governance Policy that will be overseen by the study investigators.

The database design will include appropriate physical and user level data security as well as full audit trails of access. Personal and clinical data will be separated by pseudonymization to enhance data security; the code linking personal and clinical data will only be available to the study PIs. The study database manager will be responsible for day-day management of the database, including generating quality control reports and daily query logs for resolution.

Data will be stored long-term (>20 years) for research purposes. Data access will be regulated by the study investigators who will assess written requests for access to data, to ensure that the user is an appropriate health researcher or policy maker. Data access processes will be transparent and auditable, and without unreasonable barriers to appropriate use of data. We anticipate sharing data with a range of academic, clinical, governmental and commercial organisations. The study PIs will also establish outreach to potential data users to maximise the exploitation of the data for research and policy decisions. Data sharing will follow all relevant national and international guidelines and legislation; in particular, the research data will only be shared in an anonymized format (ie all personal identifiers removed) to ensure preservation of privacy of participants.

The Chief Investigator will preserve the confidentiality of participants taking part in the study and is registered under the Data Protection Act.

15. Withdrawal process for participants

Participants may withdraw from the study at any time. Participants will be asked to submit their request to withdraw in writing to the LOLIPOP 100K study co-ordinator (contact details TBC). Reasons for withdrawal will be documented where provided. On withdrawal, the participants identifiable samples will be destroyed and linked data withdrawn firm release for further research. If samples or data have been previously used, or are anonymised, we will not be able to retrieve them.

16. Process for payments to participants

Reasonable travel expenses incurred by the participant in attending the research clinic will be reimbursed, supported by receipts where appropriate. Researchers will not receive payment over and above normal salary for recruiting the subjects in the study.

17. List of recruiting and referring sites

The North West London Primary Care Research Network have agree to support recruitment to the study. The precise details of which PCRN sites will participate in recruitment will be defined as the study gathers its regulatory approvals.

18. Training and authorising of research staff

Research staff will be trained in all study protocols according to the study SOP. Competencies will be verified at the end of training and at regular intervals thereafter. Training and delegation logs will be maintained.

19. Study management

The study investigators will guide the direction of the research, including resource allocation, developing collaborations, dissemination and pathways to impact. The investigators will be supported by the study programme management team based at Imperial College London (Ninha Silva).

20. Financial / Conflicts of Interests declarations

There are no financial or other conflicts of interests to declare. Any conflicts of interest that arise during the course of the study will be reported to the study sponsor.

21. Publication and dissemination of results

Dissemination and exploitation will include a range of communication strategies, including:

i. Project website. The website will serve as a platform for other scientific colleagues, clinicians and healthy policy makers to familiarise themselves with our research and its impact.
ii. Publications. The results of the research will be made available as an project report, submitted reports for publication in high-impact scientific journals.
iii. Public awareness. We will engage local media to disseminate knowledge of the research. We will also participate actively in local events focussed on CVD and T2D in South Asians.
iv. Awareness amongst scientific and clinical experts. We will actively prepare and submit abstracts for submission to national and international workshops/conferences to share results with other scientific colleagues, clinicians and health policy makers.

22. Intellectual Property arrangements.

Our expectation that the LOLIPOP 100K dataset will be used in a wide range of collaborative research, involving partners from a range of settings. The research team will notify the R&D office of potential intellectual property arising from the work. Arrangements for protection of

intellectual property will be made based on the nature of the discovery, and the inventive contribution of the partners involved.

23. Auditing

The study may be subject to inspection and audit by the sponsor and other regulatory bodies to ensure protocol compliance and adherence to GCP and the UK Policy Framework for Health & Social Care Research 2017). Participating NHS sites are required to comply with any requests for audit by the Sponsor or their applicable site R&D Office and ensure that the study documentation and information is available.

23. Deviations

Protocol deviations, non-compliances, or breaches from the approved protocol will be reported to the Sponsor in compliance with Sponsor processes.

24. Amendments

Approval for any subsequent changes to the study conduct, design or management will be notified to the original approving REC & HRA and any other relevant regulatory authority via the UK Amendment process (http://www.hra.nhs.uk/research-community/during-your-research-project/amendments/) Authorisation will be sought from the study sponsor for any future substantial and non-substantial amendments arising during the course of the study, prior to submission to the relevant Research Ethics Committee & HRA. Changes to the study will not be implemented until REC/HRA approval has been obtained unless the clinical need warrants this, for example when urgent safety measures are required. The Chief Investigator or designee will work with sites (R&D departments at NHS sites as well as the study delivery team) so they can put the necessary arrangements in place to implement the amendment to confirm their support for the study as amended. NHS R&D Amendment continuation of capacity and capability will be sought (where applicable) from the participating sites before any substantial changes can be implemented at the applicable site.

25. Annual Progress Reporting

An annual progress report will be submitted to the sponsor and the approving REC/HRA by the CI on every 12 months from the date of commencement of the study.

26. Ethics favourable opinion, HRA approval and NHS R&D

The Chief Investigator will obtain approval from a recognised NRES Research Ethics Committee & Health Research Authority (HRA) or the corresponding bodies within the devolved participating nations. The study will be submitted for assessment of capacity and capability at each participating NHS Trust using the HRA/REC approved study document set and Organisation Information Document (OID) & Schedule of Events (SoE). The Chief investigator or designee will prepare the documents submission to participating sites and utilise the specified e-mail templates as per the national recommended process. The Chief Investigator will require a copy of each participating site's confirmation of capacity and capability, agreed OID, PI & research teams current CV and GCP certificates, before accepting participants into the study. This information will be stored in the Main Study file located at the research team's offices at Ealing Hospital. The study will be conducted in accordance with the UK Policy Framework for Health & Social Care research . For any Non-NHS sites, a Site-Specific Assessment Form (SSI) will need to be submitted for REC review. The study will be conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

27. Indemnity

Imperial College London holds negligent harm and non-negligent harm insurance policies which apply to this study

28. Sponsor

Imperial College London will act as the main Sponsor for this study. Delegated responsibilities will be assigned to the NHS trusts taking part in this study.

Table. Power calculations for the proposed LOLIPOP 100K study. Table summarises i. the incidence of major disease endpoints in South Asians, based on the previously completed LOLIOP study (per 100,000 participants, per year), and ii. the expected number of disease cases that would occur after 10 years in the LOLIPOP 100K study of 100,000 participants. Relative risks detectable are then provided for each of the diseases in a nested case control study (2 controls per case, 90% power), under 3 experimental designs: i. single marker study ($P<0.05$), ii. Metabolome wide association study (MWAS, $P<10^{-5}$) and iii. Genome-wide association study (GWAS, $P<5x10^{-8}$). Power is calculated based on an exposure present in 20% of the population, or minor allele frequency 20% for genetic association.

Disease	Incidence	Expected cases	Single marker study	MWAS	GWAS
Death (all cause)	612	12240	1.08	1.16	1.14
Type 2 diabetes	950	19000	1.06	1.13	1.11
Cardiovascular disease	864	17280	1.06	1.14	1.11
Cancer (all cause)	625	12500	1.08	1.16	1.14
Myocardial infarction	453	9060	1.09	1.19	1.16
Stroke	221	4420	1.14	1.19	1.24
Breast cancer (F)	251	2510	1.19	1.39	1.32
Prostate cancer (M)	158	1580	1.23	1.51	1.42
Colo-rectal cancer	47	940	1.31	1.69	1.56
Pancreatic cancer	22	440	1.47	2.12	1.89
Stomach cancer	18	360	1.52	2.28	2.01
End stage renal failure	15	300	1.59	2.45	2.14

References

1. Collaboration NCDRF. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet. 2016;387(10027):1513-30.
2. collaborators NCDC. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. Lancet. 2018;392(10152):1072-88.
3. Collaboration NCDRF. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet. 2016;387(10027):1513-30.
4. Tan ST, Scott W, Panoulas V, Sehmi J, Zhang W, Scott J, et al. Coronary heart disease in Indian Asians. Glob Cardiol Sci Pract. 2014;2014(1):13-23.
5. Prabhakaran D, Jeemon P, Roy A. Cardiovascular Diseases in India: Current Epidemiology and Future Directions. Circulation. 2016;133(16):1605-20.
6. Sattar N, Gill JM. Type 2 diabetes in migrant south Asians: mechanisms, mitigation, and management. Lancet Diabetes Endocrinol. 2015;3(12):1004-16.
7. Wild SH, Fischbacher C, Brock A, Griffiths C, Bhopal R. Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003. J Public Health (Oxf). 2007;29(2):191-8.
8. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015;3(7):526-34.
9. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet. 2004;364(9438):937-52.
10. Chambers JC, Obeid OA, Refsum H, Ueland P, Hackett D, Hooper J, et al. Plasma homocysteine concentrations and risk of coronary heart disease in UK Indian Asian and European men. Lancet. 2000;355(9203):523-7.
11. Tillin T, Hughes AD, Mayet J, Whincup P, Sattar N, Forouhi NG, et al. The relationship between metabolic risk factors and incident cardiovascular disease in Europeans, South Asians, and African Caribbeans: SABRE (Southall and Brent Revisited) -- a prospective population-based study. J Am Coll Cardiol. 2013;61(17):1777-86.
12. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S1-45.
13. Gopal DP, Usher-Smith JA. Cardiovascular risk models for South Asian populations: a systematic review. Int J Public Health. 2016;61(5):525-34.
14. Bhopal R, Fischbacher C, Vartiainen E, Unwin N, White M, Alberti G. Predicted and observed cardiovascular disease in South Asians: application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data. J Public Health (Oxf). 2005;27(1):93-100.
15. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011;43(10):984-9.
16. Warren HR, Evangelou E, Cabrera CP, Gao H, Ren M, Mifsud B, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet. 2017;49(3):403-15.
17. Iotchkova V, Huang J, Morris JA, Jain D, Barbieri C, Walter K, et al. Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. Nat Genet. 2016;48(11):1303-12.
18. Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet. 2015;47(11):1282-93.
19. Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. Nat Genet. 2008;40(6):716-8.

20. Chambers JC, Abbott J, Zhang W, Turro E, Scott WR, Tan ST, et al. The South Asian genome. PLoS One. 2014;9(8):e102645.
21. Chambers JC, Zhang W, Zabaneh D, Sehmi J, Jain P, McCarthy MI, et al. Common genetic variation near melatonin receptor MTNR1B contributes to raised plasma glucose and increased risk of type 2 diabetes among Indian Asians and European Caucasians. Diabetes. 2009;58(11):2703-8.
22. Chambers JC, Zhao J, Terracciano CM, Bezzina CR, Zhang W, Kaba R, et al. Genetic variation in SCN10A influences cardiac conduction. Nat Genet. 2010;42(2):149-52.
23. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541(7635):81-6.