NIHR Global Health Research Unit for Cardiovascular Disease and Type-2 Diabetes in South Asians at Imperial College London

Project title: Understanding the patterns and determinants of health in South Asian people - South Asia Biobank.

Unit theme: Strengthening surveillance of Cardiovascular Disease, Type-2 Diabetes and their risk factors in South Asia.

Protocol - version 4.
8th September 2022

Investigators

Investigator name	Institution	Country
Malay Mridha	BRAC University	Bangladesh
Ananya Gupta	Devki Devi Foundation	India
John Chambers	Imperial College London	United Kingdom
Sophie Day	Imperial College London	United Kingdom
Gary Frost	Imperial College London	United Kingdom
Jaspal Kooner	Imperial College London	United Kingdom
Marisa Miraldo	Imperial College London	United Kingdom
Nick Oliver	Imperial College London	United Kingdom
Franco Sassi	Imperial College London	United Kingdom
Jonathan Valabhji	Imperial College London	United Kingdom
Anjana Ranjit Mohan	Madras Diabetes Research Foundation	India
V Mohan	Madras Diabetes Research Foundation	India
Guha Pradeepa	Madras Diabetes Research Foundation	India
Khadija Irfan Khawaja	Services Institute of Medical Science	Pakistan
Prasad Katulanda	University of Colombo	Sri Lanka
Anuradhani Kasturiratne	University of Kelaniya	Sri Lanka

Background

Importance of cardiovascular disease and type-2 diabetes in South Asians

Cardiovascular disease (CVD) and type-2 diabetes (T2D) are leading, closely interlinked global health challenges: 422M people live with T2D, while CVD caused 17.5M deaths in 2012 (31% of global mortality).[1,2] The burden of T2D and CVD are especially high in South Asia, the most populous and densely populated region of the world (1.73 billion people in 2015, 25% of world population). India alone is home to ~74M people with T2D.[3] Furthermore, the incidence of CVD is rising sharply across South Asian populations, in direct contrast to the falling CVD mortality rates observed in the US and most Western European countries.[4] T2D and CVD pose enormous economic burdens on individuals, families and healthcare systems, and thus make a major contribution to poverty, inequality and social instability across the South Asia region.[5] The World Health Organisation, United Nations and the International Diabetes Federation recognise that there is an urgent need for action to reverse the current epidemics of CVD and T2D in South Asians.

Importance of understanding risk relationships for NCD prevention

Accurate knowledge of the relationships between clinical and laboratory measures of metabolic and cardiovascular risk, with future incidence of T2D and CVD is an essential component of population based strategies for prevention of NCDs.[6] Appreciation of risk relationships raises awareness of disease determinants, guides appropriate pharmacological and lifestyle intervention, and informs public health policy and rational resource allocation. Identification of the primary risk factors for CVD and T2D in the population, and the implementation of systematic approaches to NCD risk stratification and prevention (including through use of multivariate risk prediction tools), have contributed to the ~70% reduction in incidence of CVD documented in Western Europe and USA over the last two decades.[7] However national mortality statistics from the UK show that health gains in UK South Asians have been substantially smaller than amongst the UK white population, suggesting that current approaches to risk stratification and intervention are less effective and require further optimisation for the South Asian ethnic group.[8]

Risk factors for T2D and CVD in South Asian populations

To date, few studies have assessed the incidence of T2D and CVD, or the relationship of these diseases to baseline risk factors in South Asian populations. Further, the available, completed investigations have all limited by small sample size and focus on a narrow range of geographic settings (typically single city).[9,10] Although the WHO and others have proposed tools for prediction of CVD amongst South Asian populations, these tools have major limitations. In particular, risk prediction tools such as GLOBORISK are derived based on: i. risk relationships observed in exclusively European and North American populations, ii. the prevalence of CVD risk factors amongst South Asian derived from small-scale population surveys, and iii. WHO estimates for country specific CVD mortality rates derived from verbal autopsy studies (the Sample Registration System, comprising ~6M across ~1M households).[11] The validity of these assumptions is uncertain, and may explain the paradoxical observation that GLOBORISK predicts young South Asians to be at higher risk of CVD compared to elderly South Asians, for the same risk factor burden. As a result of these limitations, there is a deficiency of knowledge to inform approaches for accurate quantification of risk for T2D and CVD amongst South Asians, that can be used to inform national or state healthcare policy amongst the ~1.7 billion South Asians on the Indian subcontinent.

Existing frameworks for monitoring CVD and T2D in South Asia

Accurate knowledge concerning the burden and future trajectories of CVD and T2D, is also central to the delivery of effective healthcare interventions and of health policy for prevention and control of these non-communicable diseases (NCDs). The information delivered by NCD surveillance enables prioritisation of resource allocation, as well as evaluation of healthcare and health system interventions for impact, including clinical impact and cost-effectiveness.

Current systems for surveillance of NCDs such as CVD and T2D in South Asia, are fragmented and incomplete.[12] For example, in Pakistan the WHO STEPS survey of NCDs carried out in 2014 included

just 7,700 people from 2 states, relied on self-reported disease, did not collect biological sample for diagnosis of T2D, and made no assessment of quality of care for T2D or CVD.[13] Similar limitations apply to current surveillance data for Bangladesh and Sri Lanka. In India, the INDIAB study of T2D burden and control amongst ~57,000 people from 15 states represents a step change for surveillance of T2D in the region, and provides valuable insights into the heterogeneity in disease risk that exist even within one country.[14] However, the cross-sectional design of INDIAB, and the lack of data for CVD, limit utilization of these data for the wider impact for NCD healthcare and health system interventions. The absence of nationally representative data concerning the burden of NCDs, and behavioral and clinical risk factors in South Asia, is a major challenge in the fight against NCDs in the region, and contributes to widen global health inequality in South Asia.[15]

There is a major need for improved understanding of the burden of T2D and CVD across South Asia, and of the underlying determinants of these major diseases, to underpin both targeted and population-based approaches for disease prevention, in South Asians.

Objectives

This project represents a collaboration between academic and clinical experts in NCD prevention and control from leading institutions in Bangladesh, India, Pakistan, Sri Lanka and the UK. With initial 4 year funding from the UK National Institutes for Health Research, we will pursue a program of translational research in South Asian that aims to:

- Strengthen NCD surveillance systems in Bangladesh, India, Pakistan and Sri Lanka, by establishing a network of >200 surveillance sites across these South Asian countries.
- Using standardised approaches, complete structured health assessments on a representative sample of ~170,000 South Asian men and women aged 18+ years, living at the >200 surveillance sites.
- Collect and store additional biological samples, to create South Asia Biobank as a resource of clinical data and biological samples, for future epidemiological research into the health conditions of importance in South Asia.
- *In the longer-term*, continue longitudinal follow-up of the 170,000 people screened, to ascertain health outcomes, including fatal and non-fatal NCD outcomes. Thereby enable quantification risk relationships, including identification and evaluation of biomarkers, underlying T2D, CVD and other chronic conditions.
- In addition use a subset of the surveillance sites as Health observatories for evaluation of healthcare interventions targeted at improving prevention and control of T2D and CVD in South Asia.
- Thereby provide the first robust, nationally representative data on the burden of T2D, CVD and their risk factors, and on the quality of care for these major NCDs to quantify needs, in South Asian populations.
- Use South Asia Biobank data and samples, to pursue a wide range of epidemiological research into the genetic, environmental and lifestyle factors influencing health and disease, including T2D and CVD in South Asian populations.

Methods

Study design

We will establish a network of community based, NCD surveillance sites in Bangladesh, India North, India South, Pakistan and Sri Lanka (N~40 per geographic region; N~200 in total). Surveillance sites will be centred on primary, community healthcare units, and distributed in representative rural and urban settings. At each of the ~200 surveillance sites, we will carry out systematic, structured assessments of ~750 people from the local, resident adult population (age 18+ yrs), to quantify i. the burden of CVD and T2D, and their risk factors, and ii. the quality of preventative care and disease management for CVD and T2D in the population. Data collection is specifically designed to provide quantitative data on the core NCD surveillance indicators in the WHO Global Monitoring Framework (**Appendix 1**).[16]

Data will be collected using shared methodology and standardised operating procedures, and will include measurement of blood based biomarkers for T2D and CVD. Results will be captured electronically to provide a rich and robust resource on ~170,000 men and women from across South

Asia, and made available to the health policy makers and other stakeholders to enable assessment of the needs and priorities of the population for NCD prevention and control. In addition biological samples (whole blood, serum and urine) will be collected from the 170,000 people screened to provide a unique and powerful resource ("**South Asia Biobank**") for high-quality epidemiological research into the genetic, environmental and lifestyle factors influencing and disease, including T2D and CVD in South Asians.

We will continue to monitor the health of South Asians in our surveillance sites long term (>20 years, pending funding), to i. define the prospective relationships between risk factors, biomarkers and incident NCDs, and ii. to document the secular changes in the prevalence of NCD risk factors, and the burden and management of NCDs in the population. The surveillance sites will thus additionally form the backbone of a health observatory system that will additional enable the impacts of health systems and community interventions focused on NCD prevention and control, to be assessed.

Selection of Surveillance sites

We will study at least 40 surveillance sites in each of the 5 study regions (Bangladesh, India North, India South, Pakistan, Sri Lanka; N>200 sites in total). Selection of sites will be done in such a way that we are able to generate credible estimates on burden of CVD and T2D, and their risk factors from a sample that are representative of the country, rural and urban areas separately and for each of the major administrative regions. In outline, site selection for each country/region will be as follows:

Bangladesh. We will select 42 sites, distributed across the entire country (**Appendix 2**). Bangladesh has 8 divisions, with substantial variation in population size between divisions. In addition, the Chittagong Hill Tracts (CHT), which comprise 3 districts, are different from the rest of Bangladesh with respect to population density and ethnicity. We will select 40 surveillance sites from the 8 divisions (excluding the CHT), with the number of surveillance sites in each division being proportionate to the population size of the division. In addition we will chose a further 2 surveillance sites from CHT (thus N=42 total). The number of rural and urban sites within a division will be proportional to the rural and urban population with a condition that there will be at least 1 urban site in each of the division. The CHT will have 2 rural sites.

The surveillance sites will be selected using a multistage sampling approach. For the selection of rural sites we will randomly select one district from each division and the CHT. We will then randomly select one sub-district from the district/CHT. Afterwards, one or more community clinics (CC) will be selected from the respective sub-district depending on the number of rural sites within the district. The ward in which the CC is located will be our rural surveillance site. For the selection of urban sites, one urban dispensary (UD)/CC will be randomly selected from each of the 12 city corporations. The *mahalla* in which the UD/CC is located will be the urban surveillance site. If the ward/mahalla is too large and has more than 2000 population, we will divide the ward/mahalla into segments. The segment that has the CC/UD will be our surveillance site. Everyone who meets the inclusion criteria within the selected site will be recruited in the surveillance.

India North. We will focus on 40 surveillance sites located in the South Delhi Municipal Corporation (SDMC), in the Delhi National Capital Region (NCR). SDMC serves a population of ~10M people, with diverse demographics characteristics, and includes both urban and rural villages. The South and Najafgrah zones of SDMC will be classified as one hub, while the Central and West zone will be classified as the second hub. We will compile a list of the basic health care units in each SDMC zone (Appendix 2), and classify each as urban / rural based on governmental data. If a health care unit caters to a large population (>50,000), the area will be divided into smaller colonies of 10-50,000 people. From this compiled list of SDMC locations, we will then select 40 surveillance sites at random, but including urban and rural locations in a 1:1 ratio. The population catered to by the health care units will be mobilised by the ASHA workers affiliated to the established surveillance sites.

India South. We will conduct surveillance in both urban and rural areas. The surveillance sites will be villages in rural areas and census enumeration blocks (CEBs) in urban areas. The ultimate stage units

will be households in both areas. The rural component of the surveillance is planned to be conducted in 25 selected villages in Cheyur taluk (Appendix 2), which is situated in Kancheepuram district of Tamil Nadu state in southern India. All individuals who meet the inclusion criteria in each household within the selected villages will be invited to participant in the surveillance (n=20,000).

The urban component of the surveillance will be carried out in Chennai, which has a population of 10,705,000 individuals. Chennai is divided into 15 zones and 200 wards by the Chennai Corporation. Of the 200 wards, 40 wards will be randomly selected to represent all 15 zones of the Corporation of Chennai. From the selected 40 wards, 10 Census Enumeration Blocks (CEBs) will be randomly selected from each ward (n=400). The next stage of sampling will be based on systematic sampling where 50 households from each CEB will be selected with a random start. One individual will be selected using the WHO KISH method and meets the inclusion criteria in each household will be invited to participant in the surveillance (n=20,000) (Appendix 2).

Pakistan. Pakistan has a population of ~214M people, and is divided administratively into four provinces (Balochistan, Punjab, Sindh, Khyber Paktunkhwa) and the Islamabad Capital Territory (ICT). Each province is further divided into divisions, towns and union councils. The union council is the smallest division, and typically consists of ~30,000 people. Each union council has basic health unit and outreach staff. The union council will be our sampling unit for surveillance.

We will focus the surveillance project on the Punjab Province, the most populated state of the country (N=110M). We will aim to establish surveillance sites in each of the 36 states that comprise the Punjab Province, distributed between in proportion to population size, with approximately 1 site per 3 million population. Surveillance sites will be selected using a multistage sampling approach. During the first stage, target towns will be selected for each province, based on population size of the province (one town per 5-6 sites). In each town, the available administrative lists of union councils will be used to select rural and urban union councils in a 1:1 ratio. As stage 1, surveillance will start in central Punjab (Appendix 2), and will be extended to North and South Punjab on successful completion of stage 1.

Sri Lanka. Sri Lanka has 9 Provinces, that are in turn divided administratively into 25 Districts, and then into further administrative units of which the smallest is the Grama Niladhari (GN). We will select 100 surveillance sites from all the districts and provinces. Each site will have 500 persons eligible to be a participant of the surveillance. The number of surveillance sites for each province will be proportionate to the population size of the province. The number of sites for each district within a province will again be proportionate to the population size of the district. Within a district we will allocate the surveillance sites based on the proportion of rural and urban population. The provisional allocation of surveillance sites by district and province is summarized in Appendix 2.

For the selection of rural and urban sites within a district, we will make a list of rural and urban Grama Niladhari (GN) divisions having a Primary Medical Care Units (PMCU) or a Divisional Hospital with a healthy lifestyle center. From the list of these GNs we will randomly select the number of surveillance sites needed for the district. If any GN is too large and has more than 1000 population, we will divide the GN into segments based on already existing demarcations in the electoral lists. The segment that has the PMCU/Divisional Hospitals will be our surveillance site. Everyone who meets the inclusion criteria within the selected site will be invited to be a participant of the surveillance.

Participants: entry criteria and sample size

All adults South Asian men and women living within the geographic boundaries of the selected surveillance site (as described by the relevant national administrative systems), will be invited to take part in the research. Recruitment targets (participants with appropriate consent, data and samples collected) by region are provided in **Table 1**.

Inclusion criteria will be:

1. Age 18+ years
2. Male or Female
3. South Asian ancestry

4. Permanent resident of the surveillance site (resident for >12 months)

Exclusion criteria will be:
1. Planning to leave the study site within the next 12 months
2. Cancer or other serious illness expected to reduce life expectancy to less than 12 months
3. Unable or unwilling to give consent

Table 1. Recruitment targets by country / region for the Surveillance study.

Region	**Recruitment target**	**Partner organisations**
Bangladesh	50,000	BRAC University, Dhaka
North India	30,000	DDF (Max Healthcare), Delhi
South India	40,000	Madras Diabetes Research Foundation, Chennai
Pakistan	30,000	Service Institute of Medical Science, Lahore Punjab Institute of Cardiology, Lahore King Edward Medical University, Lahore
Sri Lanka	50,000	University of Colombo, Colombo University of Kelaniya, Ragama

Recruitment

At each study site, governmental census data and other available household listings will used, supported by house-house visits, to fully enumerate and identify the resident population. House-to-house visits will be done by the local primary care worker, in partnership with the research team. The purpose and potential benefits (personal and societal) of the surveillance will be explained. The research team will then obtain demographic details for the adult population of each household. People eligible and consenting to take part in the surveillance study will be asked to attend a nearby health facility in the fasting state, for the surveillance health assessment. Explanations of the project will be provided in writing and using videos, available in relevant South Asian languages, and supported by bilingual translators. We will also engage the community elders (eg teachers, employers, religious leaders) as trusted third parties, to support and facilitate engagement of the community in the study. Based on previous experience with population studies in these communities, participation rates are expected to exceed >70%; the availability of household enumeration data will enable participation to be monitored in detail, including by age, gender and major socio-economic groups.

Study procedures

Participants will be asked to attend in the morning after an overnight fast (water only after midnight). All participants will complete a structured assessment by trained members the dedicated research team, in five complementary domains: i. Registration and consent; ii. Questionnaires; iii. Physical measurements; iv. Biological samples; v. Physical activity monitoring and vi. Clinical reporting. All study procedures implemented represent well established tools, that are validated for assessment of NCDs in population studies. The procedures will be carried out by trained research staff, using equipment and protocols that are standardized between countries and surveillance sites.

i) **Registration and consent**. Written, informed consent will be obtained from all participants for data collection, and inclusion in the research. Informed consent will include permission for the data and samples collected to be used for NCD 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. Consent will be facilitated using videos (available in major South Asian languages) and supported by bilingual translators. A unique study ID will be allocated to each participant.

ii) **Questionnaires**. An interviewer administered health and lifestyle questionnaire will be used to collect information on behavioral risk factors (smoking, alcohol habit, physical activity and fruits/vegetables consumption), personal and family medical history, medications, socio-economic status and knowledge of NCDs. The questionnaire is founded on the extended WHO STEPS questionnaire that is widely used in global NCD surveillance,[17] but adapted for use in South Asia, including through incorporation of additional questions. In addition, participants will be asked to complete a 24 hour dietary recall to enable quantification and comparison of nutrient intakes. The interviewer administered questionnaires will be translated into locally relevant languages, and supported by graphics representations to facilitate collection of high quality data.

iii) **Physical measurements**. Including: a) Anthropometry (height, weight, waist and hip circumference, and bio-impedance for body fat composition); b) Blood pressure by digital device; c) Cardiac evaluation by 12 lead ECG to identify arrhythmia, LVH and previous myocardial infarction; d) Retinal photography for assessment of retinal disease, including hypertensive and diabetic retinopathy; and e). Respiratory evaluation by spirometry to assess for smoking/environment-related lung injury.

iv) **Biological samples**. We will collect venous blood samples (~25mls, by venesection using trained phlebotomists) for measurement of fasting glucose, lipid profile and HbA1c and other clinical relevant markers of cardiovascular health. This will comprise collection of blood into EDTA, serum and citrate vacutainer tubes, and into tubes designed for RNA preservation. An Oral Glucose Tolerance Test will be carried out in a subset of participants, enabling validation of diabetes classification by HbA1c. A spot urine sample (10ml) will also be collected for analysis of albuminuria and other biomarkers. Laboratory assays will be carried out using validated near-patient assays, and aliquots stored (-80C) for both external QC and future molecular epidemiological research.

v) **Physical activity monitoring**. This will be done using the Axivity device, worn on the wrist for 7 days. This is a light-weight device, that is shaped and worn like a small wrist-watch. The device uses triaxial accelerometry to infer participant movement and thus physical activity. The device is battery powered, waterproof and is comfortable for continuous wear. The Axivity device has recently been used to measure physical activity patterns amongst 100,000 people in the UK Biobank study.

vi) **Clinical reporting**. All results will be reviewed for clinically significant findings by a medically qualified member of the research team, who will also be responsible for ensuring that participants identified to have significant health conditions (e.g. T2D, hypertension) are referred to an appropriate facility for counseling and treatment. All study participants will receive a clinical report detailing the results of their health assessment, and the opportunity to discuss their results with the medical practitioner if significant abnormalities are present. To facilitate this, the clinical report will be made available to local community healthcare teams. Results will also be accompanied by a booklet of explanation, as well as access to an explanatory video.

Study team and visit organisation

We anticipate operating at least 2 surveillance teams per country / region (ie at least 10 teams total); the surveillance teams will rotate through the selected surveillance sites over the course of study. An expedited approach with greater initial effort may be used if appropriate, to achieve the same objective.

The typical composition (human resource) of the surveillance team is summarized in **Table 2**. The study equipment that will be used in summarized in **Table 3**. Each surveillance team will work in a station based approach to maximise efficiency (**Figure 1**). The estimated duration of each station is shown; the surveillance health assessment is anticipated to take ~90 mins per person. We anticipate that each surveillance team will assess 25-40 participants per day; we thus anticipate screening 250-300 people per day across all sites (ie ~60,000 people per year, ~170,000 by Q1 2021).

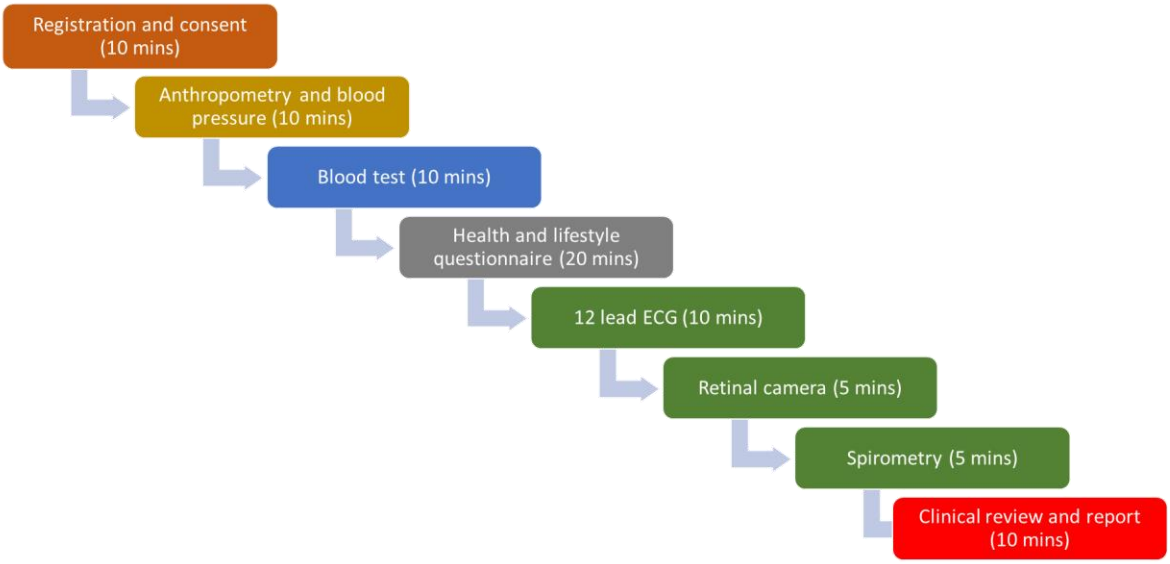
Table 2. Typical composition of each surveillance team. Additional staff will be available for transportation and security, as required in respective local contexts.

Type of staff	Number	Job responsibility
Medical officer	1	● Overall responsibility for the team ● Report on the clinical and laboratory data collected ● Clinical guidance and counseling of participants
Co-ordinator	1	● Co-ordinate the team and oversee data collection ● Implement quality control protocols
Research Assistants / nurses	5	● Registration and consent (1 Pax) ● Administer participant questionnaires (2 tables → 2 Pax) ● Anthropometry and blood pressure (1 Pax) ● Assist Medical officer in clinical review (1 Pax) ● Carry out household listing (afternoon activity, all)
Phlebotomist	1	● Collection of blood samples and urine samples
Laboratory technicians	4	● 12 lead ECG (1 Pax) ● Retinal photography (1 Pax) ● Spirometry (1 Pax) ● Analyze blood samples (1 Pax) ● Process biological samples for storage (afternoon, all)

Table 3. Provisional equipment list for the surveillance project.

Measurement	Device	Number
1. Core equipment (NCD "toolkit"; one toolkit per site)		
Height	Seca Wall Mounted Tape Measure	1
Weight	Salter Electronic Scale	1
Tape measure	Plastic tape measure	1
Blood pressure	Omron device or Aneroid sphygmomanometer + stethoscope	1
Glucometer	To be selected as a device in routine local use	1
Tablet	7" Fusion5 Android Tablet or equivalent	1
2. Special equipment (one set per surveillance team)		
Weight & bioimpedance	OMRON BF511 (clinically validated)	2
Blood pressure	Omron M3 IT or equivalent	2
Spirometry	MIR: Minispir or Spirobank Smart	1
12 lead ECG	GE: Mac800 or equivalent	1
Retinal photography	Remedio: Fundus on a Phone or equivalent	1
Physical activity	Axivity device	~200
Blood tests: HbA1c	SD Biosensor: Multicare HbA1c	1
Blood tests: Lipids	SD Biosensor: Lipidocare	1
Tablets	7" Fusion5 Android Tablet or equivalent	15
Centrifuge	Remi C854/6	2
Fridge	Local vendor, 4C	1

Figure 1. Station based approach to NCD surveillance.



Sample storage

We will collect ~25mls of blood and 10mls of urine from participants during the surveillance exercise. Aliquots of whole blood, serum, plasma, and urine will be stored at -80C for future quantification of biomarkers for cardiovascular and metabolic health (eg measurement of NTpro-BNP as a measure of heart function, or insulin/C-peptide for glucose metabolism), 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 use 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).

Follow-up and future recontact

Ambition

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 Asian populations. There are two broad approaches to follow-up that will be adopted:

- Routine record linkage. In Europe, US and many part of Asia it is now possible to use identify participants health outcomes from their locally or nationally held health records. Examples include death notification, cancer registries, hospitalization, results of laboratory investigations and imaging. 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. The written consent from all participants at the baseline visit will include 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.

Follow-up visit 1: Identify COVID-19 incidence and outcomes in study participants

South Asian people are at high risk both for infection with SARS-CoV2 virus, and for developing severe or fatal COVID-19. Addressing this inequality is a major public health priority. We propose to use our surveillance study cohort to improve understanding of the impact of COVID-19 on South Asian populations. We will identify incident COVID-19, amongst South Asian men and women who are participants of our prospective population study. Participants will undergo testing for SARS-CoV2, and complete a validated questionnaire to assess symptoms, relevant behaviours, attitudes to vaccination, and adverse outcomes. We will use our available comprehensive baseline phenotypic data, including molecular characterisation and stored biological samples, collected before the COVID-19 pandemic, to determine what are the major risk factors for infection with SARS-CoV2, severe COVID-19, or prolonged COVID-19, amongst South Asians. We will focus initially on recognised risk factors such as adiposity, raised blood pressure, diabetes, cardiovascular disease, health behaviours, socio-economic indicators, and biochemical measures. We will compare our results in South Asians with equivalent data for Europeans available from other studies, to determine whether these known risk factors explain the high risk of COVID-19 in Asians. Our research will thereby determine the reasons underlying the high burden of COVID-19 in South Asians and generate knowledge that will inform health policy and practice for prevention and control of the disease amongst South Asians.

Methods will comprise a single follow-up visit lasting ~30 minutes during which the study team will ask the participant to complete the following low-risk assessment:

i) Consent. Informed consent will be obtained and will include permission for the data and samples collected to be used for COVID-19 research, including data sharing with national and international bodies concerned with prevention and control of COVID-19.

ii) Questionnaire. An interviewer administered questionnaire will be used to collect information on i. COVID-19 symptoms (acute and prolonged); ii. hospital admission for COVID-19; iii. results of COVID-19 diagnostic tests and vaccines uptake; iv. health-related and social impact due to COVID-19; v. vaccine hesitancy; and vi. mood and financial well-being. Questionnaires are based on those developed by the REACT2 study and in collaboration with Wellcome Trust, for use in HIC and LMIC settings. The questionnaire will be interviewer administered. In addition, verbal autopsy is carried out using validated tools to assess probable cause of death for people who have died.

iii) Physical measurements. These include: a) Anthropometry (weight, waist and hip circumference); and b) Blood pressure by digital device.

iv) Biological samples. Venous blood sample (~8mls, by venesection using trained phlebotomists) will be collected for serological testing for SARS-Cov-2 antibodies, as well as for other laboratory assays relevant to COVID outcomes (e.g., cardiometabolic, metabolomic and proteomic assays).

This sub-study is funded through an award from the UK Medical Research Council (**MR/V040049/1**; CoV-Ind-UK: Prospective investigation of the determinants for COVID-19 outcomes amongst South Asians in India and the United Kingdom. All assessment approaches will follow national guidance in relation to safe distancing and related measures for COVID-19 prevention.

Quality control and data analysis

Quality control

The surveillance teams, comprising research assistants, laboratory technicians, physicians and co-ordinators will be trained to follow standardized protocols in each country/region. Their training modules will include interviewing techniques, ethics and specific instructions for each of data (demographic, socio-economic, food security, behavioral risk factors, medication and lifestyle practices, physical measurement and collection of biological samples etc.). Revalidation of the research teams in study procedures will be done at regular intervals during the study to ensure continued high quality data

collection that is harmonized across sites. Standardized operating procedures will be established for all data collection procedures.

Questionnaires will be translated into the languages appropriate to the communities, and back-translated to ensure adequacy of translation. Equipment used for physical and biological measurement will be regularly calibrated using appropriate controls/standards. The data management team will review the data collected and provide daily feedbacks on inaccuracy and missing data, and will adopt data-driven approaches to monitoring data quality (eg checks on biases in data entry, logical inconsistencies, internal correlations, digit preference, measurement drift or bias between machines and observers). If necessary, we will also consider re-assessing a random subset comprising up to 2% of the study participants, and/or re-testing a subset of biological samples to provide additional quality control data. To the extent possible, an attempt will be made to use data collection methods that are "field-friendly" (e.g., devices that are portable and can also give immediate readings and do not require additional laboratory analysis), culturally-acceptable and minimally-invasive in order to reduce subject attrition and improve logistical feasibility.

Data analysis

The key study indicators for the surveillance component of the project are based on the WHO NCD Global Monitoring Framework (Appendix 1). Therefore, we will follow the operational definition of the indicators, method of estimation, and guidelines for data disaggregation (by sex/socio-demographic stratifiers) proposed in the framework. As we will use a multistage sampling approach, appropriate sample weight will be estimated and used during the descriptive analysis.

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). Regional differences within and between countries will also be explored. Statistical significance will be inferred at $P<0.05$, using techniques (eg Bonferroni) to control for multiple comparisons where appropriate. Power estimates for analyses of cross-sectional and longitudinal outcomes data are summarized in **Tables 4 and 5**.

Table 4. Power calculations for cross-sectional analyses. Results are expressed as i. the precision with which the prevalence for a clinical outcome of interest can be quantified and ii. the odds ratios (OR) detectable for association between a clinical outcome and an underlying exposure (exposure prevalence 20%, power 90%, P<0.05). Precision and power estimates are provide overall (N~170,000 people), at the level of the country (M~30,000 people), or at the level of the site (N~750 people). The table also summarized precision and power for low (~1%), intermediate (5-10%) and high frequency (~20-30%) outcomes of interest.

		Precision			OR detectable	
Outcome measure	**Expected prevalence**	**Overall**	**Country**	**Site**	**Overall**	**Country**
N=		170,000	30,000	750	170,000	30,000
Hypertension	30%	0.2%	0.5%	3.3%	1.04	1.10
Diabetes	20%	0.2%	0.5%	2.9%	1.05	1.12
Chronic kidney disease	10%	0.1%	0.3%	2.2%	1.07	1.16
Cardiovascular disease	5%	0.1%	0.3%	1.6%	1.09	1.23
Severe CKD	2%	0.1%	0.2%	1.0%	1.14	1.37
Stroke	1%	0.0%	0.1%	0.7%	1.20	1.54

Table 5. Power calculations for longitudinal outcomes. Table summarises the expected incidence of disease endpoints in South Asians (per 100,000 participants/year, based on follow-up of South Asians in the LOLIPOP study [2003-2017]), and thus the expected number of disease cases that occur after 5 and 10 years in the population sample of up to 170,000 South Asians proposed in the current surveillance project. Relative risks detectable are provided for a range of diseases in a nested case control study (2 controls per case, 90% power), under 3 experimental designs: i. single marker study (SMS, 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 (per 100K year)**	**Expected cases**		**RR: SMS**		**RR: MWAS**		**RR: GWAS**	
		5 yrs	**10 yrs**	**5 yrs**	**10 yrs**	**5 yrs**	**10 yrs**	**5 yrs**	**10 yrs**
Type 2 diabetes	950	8550	17100	1.09	1.06	1.19	1.13	1.17	1.11
CVD	864	7776	15552	1.09	1.06	1.20	1.14	1.17	1.11
Cancer (all cause)	625	5625	11250	1.11	1.08	1.24	1.16	1.20	1.14
Myocardial infarction	453	4077	8154	1.14	1.09	1.28	1.19	1.23	1.16
Stroke	221	1989	3978	1.19	1.14	1.42	1.19	1.34	1.24
Breast cancer (F)	251	2259	4518	1.26	1.19	1.58	1.39	1.47	1.32
Prostate cancer (M)	158	1422	2844	1.34	1.23	1.77	1.51	1.62	1.42
Colo-rectal cancer	47	423	846	1.45	1.31	2.07	1.69	1.85	1.56
Pancreatic cancer	22	198	396	1.71	1.47	2.81	2.12	2.42	1.89
ESRF	15	135	270	1.89	1.59	3.43	2.45	2.88	2.14

Study and data management

Study management

The study Steering Committee will comprise the Chief Investigator, Lead Investigators from each country/region, and senior investigators from Imperial College London (based on domain specific expertise). The Steering Committee will guide the direction of the research, including resource allocation, developing collaborations, dissemination and pathways to make an impact. Steering Committee recommendations are based on consensus, with majority voting where required. Steering committee meetings are held monthly by Skype, supplemented by face-face meetings every 6 months. Dates and venues for monthly meetings have been agreed for 2018, and for face-face meetings through to 2021. The Unit secretariat have taken responsibility for co-ordination of meetings, and for maintaining accurate minutes. The work of the study team is further supported by an International Scientific Advisory Board that will meet once a year to provide independent review and advice on the program of work.

The investigators will work closely with national leads in the partner countries to ensure project delivery and appropriate data. Country specific advisory committees will be set up make sure that the findings from the surveillance are utilized in the national policy making, program design and reporting to the global bodies. In addition county specific management committees, will coordinate the activities of the project under guidance from the Steering Committee.

Data management and protection.

We will commission a purpose built database to facilitate i. robust collection; ii. secure storage and iii. appropriate sharing / exploitation of population health data, including reporting of clinically relevant results to individual participants. The database will be commissioned under the supervision of the Chief Investigator and respective national Principal Investigators.

Data will be captured through handheld (tablet +/- smartphone) 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. Relevant national 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 (eg HIPAA compliance).[18] 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 Steering Committee.

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 country / region specific 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 Steering Committee 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 Steering Committee 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.

Sample management

As described above, ~25mls of blood and 10mls of urine 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 NTpro-BNP as a measure of heart function, or insulin/C-peptide for glucose metabolism), 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 and urine 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. The sample resource will be managed by the study Steering Committee, using mechanisms analogous to those for data access. Given the limited, and depletable nature of the sample resource, the criteria for sample use will be more stringent aiming to ensure that the maximum scientific and translational research benefit.

Dissemination and exploitation of results

The results of the research will have far-reaching relevance beyond the partners directly involved in this project. Disseminating and exploiting the newly generated knowledge on the burden and quality of care for CVD and T2D amongst South Asians is a key project objective. Dissemination and exploitation will include a range of communication strategies, including:

i. Project website. We have an open-access web page with the objectives and design of the trial, as well as relevant results (www.ghru-southasia.org). The website will also serve as a platform for other scientific colleagues, clinicians and healthy policy makers to familiarise themselves with our research and its impact.
ii. Project reports. The full results of the trial will be made available as an open access report, freely available through the study website. This will describe all aspects of the research design, experience and results relevant to policy-makers, clinicians, scientists and communities organisations. In addition we will submit reports for publication in high-impact scientific journals.
iii. Public awareness. We will engage local media to release press reports at regular intervals during 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 our experience in surveillance and our results with other scientific colleagues, clinicians and health policy makers. We will actively involve stakeholders from both the scientific and political communities in order to disseminate the results of the trial endorsed by multiple stakeholders throughout South Asia and internationally.
v. Policy and practice. The partnership includes applicants who are actively involved in healthcare policy locally, nationally and internationally. Together these partners will work together to communicate, disseminate, maximise the impact of the research, and to embed the findings into local, national and international policy and practice.

Ethics and protection of human rights

Imperial College London will be the study Sponsor and will provide Clinical Research Insurance for partners. Approval for the research will be obtained from Institutional Review Boards in each of the participating countries, as well as in the UK.

Informed written consent will be obtained from each participant before their enrollment, facilitated by videos in relevant languages and bilingual translators. As noted above, participants identified to have significant health conditions (e.g. T2D, hypertension) will be referred to an appropriate facility for counseling and treatment.

Expected impacts and benefits

The burden of T2D and CVD is especially high in South Asia. T2D affects ~56 million people, and CVD ~62 million people in India alone. The number of people affected by T2D and CVD in South Asia will double by 2030; contrasting stable or falling rates in Europe and the US. T2D and CVD pose an

enormous economic burden on individuals, families and healthcare systems of South Asia, and contribute to increasing poverty, widening health and social inequalities and regional instability. The lack of evidence to inform healthcare policy for T2D and CVD in these environments, contributes to the continued rise of these diseases in South Asia, and to widening global inequality. T2D and CVD are identified as the leading public health challenges in all South Asian countries.

Benefits to society. Through our research, we will provide robust data on the burden and quality of care in South Asia. The information collected will inform priorities for healthcare providers and policy makers, and facilitate transformation of health systems in South Asia. Effective actions based on our findings (eg through guidelines, quality standards and high quality training) will help improve efficiency, effectiveness and quality of healthcare in South Asia. Our actions are thus directly designed to promote the health needs and welfare of the people living in South Asian countries, and thereby reduce global health inequalities, regional instability and poverty.

Benefits to individuals. As part of our surveillance exercise, we will provide structured individual reports to participants that detail the results of their health assessment, and provide clear written advice for suitable measures to improve health and well-being. Furthermore, participants will be reviewed by the study Medical Officers to provide specific interpretation, explanation and guidance concerning the personal health report, and the surveillance team will facilitate onward referral of participants identified with health risks for further appropriate treatment. We this bring direct benefit to individual participants.

Sustainability and future directions

The UK National Institute for Health Research (NIHR) has provided funding to establish and maintain the network of surveillance sites for the first three years. The lead organization in each country/region has committed to maintaining the NCD surveillance program beyond the four year interval. The Chief Investigator, Steering Committee and other scientists and clinicians involved in the study, are additionally committed to obtaining ongoing funding beyond the initial award from NIHR, to enable the late outcomes of the research (>20 years) to be determined.

References

1. WHO. Global report on diabetes. World Health Organisation, Geneva; 2016. http://www.who.int/diabetes/global-report/en/. Accessed 9 May 2018.
2. WHO. Global Health Estimates: Global Deaths by cause, age and sex, 2000–2015. World Health Organisation, Geneva; 2016. http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html. Accessed 9 May 2018.
3. IDF. International Diabetes Federation Diabetes Atlas. 8th ed. International Diabetes Federation. Belgium; 2017. http://www.diabetesatlas.org/resources/2017-atlas.html. Accessed 9 May 2018.
4. Mathers CD, Salomon JA, Ezzati M, et al. Global Burden of Disease and Risk Factors. Oxford University Press. New York; 2006.
5. WHO. Action plan for the prevention and control of noncommunicable diseases in South-East Asia, 2013–2020. World health Organisation, New Delhi; 2013.
6. Hajifathalian K, Ueda P, Lu Y, Woodward M, et al. A novel risk score to predict cardiovascular disease risk in national populations (Globorisk): a pooled analysis of prospective cohorts and health examination surveys. Lancet Diabetes Endocrinol. 2015;3:339–355.
7. Townsend N, Bhatnagar P, Wilkins E, et al. Cardiovascular disease statistics 2015. British Heart Foundation. London; 2015.
8. Barnett, A.H. et al. Type 2 diabetes and cardiovascular risk in the UK south Asian community. Diabetologia 2006;49:2234–2246.
9. Jayawardena R, Ranasinghe P, Byrne NM, et al. Prevalence and trends of the diabetes epidemic in South Asia: a systematic review and meta-analysis. BMC Public Health 2012;25;12:380.

10. Anjana RM, Ali MK, Pradeepa R, et al. The need for obtaining accurate nationwide estimates of diabetes prevalence in India— rationale for a national study on diabetes. Indian J Med Res 2011;133:369–80.
11. Mathers C, Ma Fat D, Inoue M, Rao C, Lopez AD. Counting the dead and what they died of: an assessment of the global status of cause of death data. Bulletin of the World Health Organization 2005;83:171-7.
12, 15. Dans A, Ng N, Varghese, C, et al. The rise of chronic non-communicable diseases in Southeast Asia: Time for action. Lancet 2011;377:680–689.
13. Pakistan Health Research Council. Non-communicable Diseases Risk Factors Survey Pakistan, 2014 - 2015. Pakistan Health Research Council, Islamabad; 2016.
14. Anjana RM, Deepa M, Pradeepa R, Mahanta J, Narain K, Das HK, et al. Prevalence of diabetes and prediabetes in 15 states of India: results from the ICMR-INDIAB population-based cross-sectional study. Lancet Diabetes Endocrinol. 2017;5:585–96.
16. WHO. NCD Global Monitoring Framework. Ensuring progress on noncommunicable diseases in countries. http://www.who.int/nmh/global_monitoring_framework/en. Accessed 10 May 2018.
17. World Health Organization. WHO STEPwise approach to surveillance (STEPS). http://www.who.int/chp/steps/en. Accessed 10 May 2018.
18. US Department of Health and Human Services. OCR Privacy Brief: Summary of the HIPAA Privacy Rule. Washington, DC: Office for Civil Rights, HIPAA Compliance Assistance; 2003. https://www.hhs.gov/sites/default/files/privacysummary.pdf. Accessed 10 May 2018.

Appendix 1. WHO Primary and secondary indicators for NCD prevention and control

The relevant core NCD surveillance indicators from the WHO Global Monitoring Framework, that will be directly captured in our surveillance study are as follows:

Domain	Indicators
Mortality and morbidity	● Unconditional probability of dying between ages of 30 and 70 from cardiovascular diseases, cancer, diabetes or chronic respiratory diseases
Behavioral risk factors	● Total (recorded and unrecorded) alcohol per capita consumption within a calendar year in litres of pure alcohol, as appropriate, within the national context ● Age-standardized prevalence of insufficient physically active persons aged 18+ years (defined as less than 150 minutes of moderate-intensity activity per week, or equivalent) ● Age-standardized prevalence of current tobacco use among persons aged 18+ years
Biological risk factors	● Age-standardized prevalence of raised blood pressure among persons aged 18+ years (defined as systolic blood pressure ≥140 mmHg and/or diastolic blood pressure ≥90 mmHg) and mean systolic blood pressure ● Age-standardized prevalence of raised blood glucose/diabetes among persons aged 18+ years (defined as fasting plasma glucose concentration ≥ 7.0 mmol/l (126 mg/dl) or on medication for raised blood glucose) ● Age-standardized prevalence of overweight and obesity in persons aged 18+ years (defined as body mass index ≥23 kg/m² for overweight and body mass index ≥ 28 kg/m² for obesity)
Behavioral risk factors	● Age-standardized prevalence of persons (aged 18+ years) consuming less than five total servings (400 grams) of fruit and vegetables per day
Biological risk factors	● Age-standardized prevalence of raised total cholesterol among persons aged 18+ years (defined as total cholesterol ≥5.0 mmol/l or 190 mg/dl); and mean total cholesterol concentration
Drug and counseling	● Proportion of eligible persons (defined as aged 40 years and older with a 10-year cardiovascular risk ≥30%, including those with existing cardiovascular disease) receiving drug therapy and counselling (including glycaemic control) to prevent heart attacks and strokes ● Proportion of people with hypertension receiving drug therapy and counselling ● Proportion of people with diabetes receiving drug therapy and counselling

Appendix 2. Draft of proposed surveillance sites

Table 2.1 Bangladesh Sample Design

Selection of rural sites				
Division / Region (# of Districts)	District (# of sub-districts)	Sub-district (# of community clinics)	# CC selected	Names of CC
Barisal (6)	Barguna (6)	Betagi (16)	(1)	Talgachhiya Deshantarakathi Cc –Betagi (Code:10002396)
Chittagong (8)*	Chandpur (8)	Shahrasti (23)	(6)	Pathair Cc -saharasthi (Code:10003514) Khanesar cc -saharasthi (Code:10003515) Potepur Cc-saharasti (Code:10019104) Balshid Cc - Saharasthi (Code:10019229) Tamta Cc - Shaharasthi (Code:10019230) Khampar Cc - Saharasthi (Code:10019233)
Dhaka (13)	Faridpur (9)	Charbhadrashan (9)	(8)	Moulavirchar Cc - Char Bhadrashion (Code:10005146) Char Salepur Cc - Char Bhadrashion (Code:10005150) Imatmaj Molyar Dangi Cc - Char Bhadrashion (Code:10005151) Salepur Cc - Char Bhadrashion (Code:10005152) Char Sultanpur Cc - Char Bhadrashion (Code:10005153) Gazirtek Cc - Char Bhadrashion (Code:10005154) Joydeb Sarkarer Dangi Cc - Char Bhadrashion (Code:10005155) Char Jhaukanda CC - Char Bhadrasan (Code:10023641)
Khulna (10)	Kushtia (6)	Daulatpur (52)	(3)	Chatarpara Cc -daulatpur (Code:10008650) Kalidaspur Cc -daulatpur (Code:10008651) Taraguniya Cc -daulatpur (Code:10008666)
Mymensingh (4)	Sherpur (5)	Jhenaigati (22)	(2)	Gojarikura Cc - Jhenaigati (Code:10007335) Rangamati CC - Jhenaigati (Code:10020689)
Rajshahi (8)	Natore (5)	Baraigram (38)	(4)	Rayna Vorat Cc -baraigram (Code:10010109) Diyar Garfa Cc -baraigram (Code:10010112) Vabanipur Cc -baraigram (Code:10010123) Borodeha Cc -baraigram (Code:10010131)
Rangpur (8)	Dinajpur (13)	Parbatipur (40)	(3)	Sonapukur Cc - Parbatipur (Code:10011197) Barakona Cc - Parbatipur (Code:10011198) Shingimari Darikhamar Cc - Parbatipur (Code:10011228)
Sylhet (4)	Moulvibazar (7)	Sreemangal (30)	(1)	Sindurkhan Cha Bagan Cc - Sreemangal (Code:10012766)
CHT (3)	Khagrachhari (9)	Ramgarh (11)	(2)	Sudu Karbari Para Cc - Ramgarh (Code:10004532) Baratholy CC - Ramgarh (Code:10024561)
Selection of urban sites				
Division	City corporation	UD/CC in the city corporation	N	Names of UD/CC selected
Barisal	Barisal	35	(1)	Char Karanji Cc - Barishal Sadar (Code:10002513)
Chittagong	Chittagong	9	(2)	Firojshah Colony Urban Dispensary (Code:10000763) Gausul Azam Urban Dispensary (Code:10000764)
Dhaka	Dhaka	17	(4)	Johnson Road Urban Dispensary (Code:10000040) Mirpur Old Colony Urban Dispensary (Code:10000043) Motijheel Urban Dispensary (Code:10000046) Nayatola Urban Dispensary (Code:10000048)
Khulna	Khulna	1	(1)	Khalishpur Urban Dispensary, Khulna. (Code:10001803)
Mymensingh	Mymensingh	55	(1)	Char Gabindapur Cc - Mymensingh Sadar (Code:10006615)
Rajshahi	Rajshahi	3	(1)	Seruil Urban Dispensary (Code:10001566)
Rangpur	Rangpur	4	(1)	Chanderhat Jagadispur Cc - Rangpur Sadar (Code:10012296)
Sylhet	Sylhet	1	(1)	M. C. College Dispensary, Sylhet (Code:10017206)

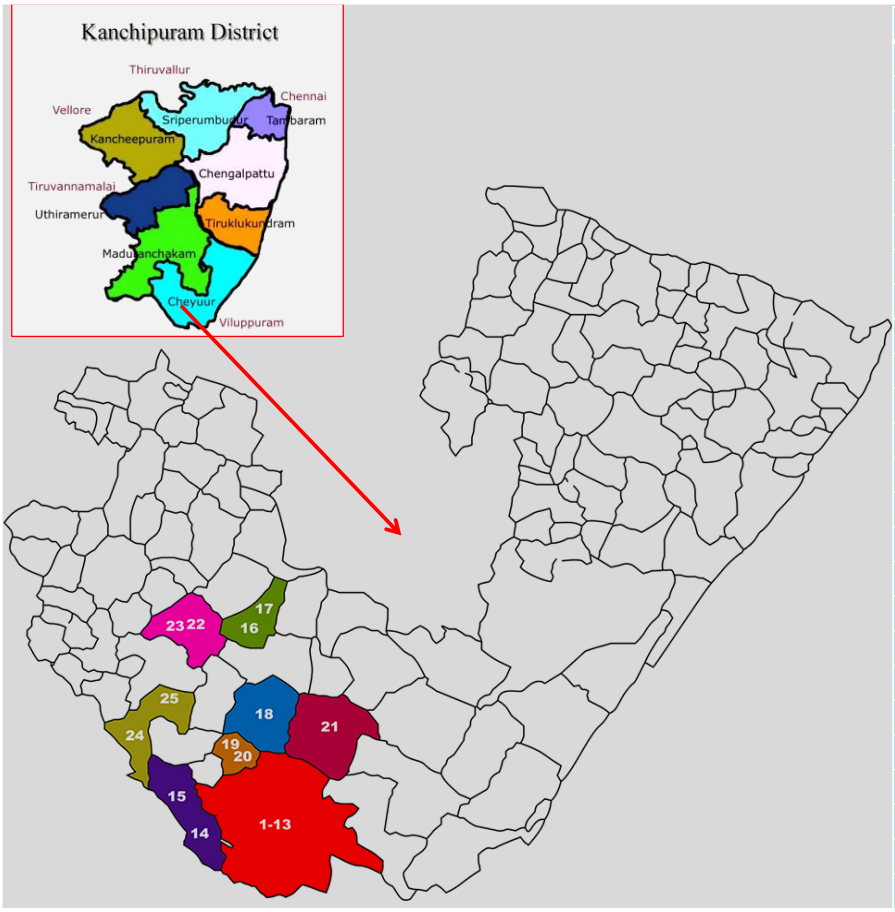
* Chittagong has 11 districts; 3 of them are part of CHT

Table 2.2. North India sample design. Delhi Municipal Corporation Surveillance Sites, New Delhi

Zone (# of UPHCs)	UPHC (# of colonies)	# of areas	Name Of Areas
Central (9)	Lajpat Nagar (8)	4	Nehru Nagar Lajpat Nagar-II Ashram West Hari Nagar Lajpat Nagar-1
Central (9)	Kalkaji (12)	6	Janta Quarters Dda Quarters Giri Nagar Govindpuri Kalkaji Colony Tara Apartment
South (11)	Munirka (18)	6	IIT Huts Vasant Vihar Huts Kusum Pahari Huts Jiya Sarai Basant Nagar Village Mochi Bagh Village
South (11)	Mehrauli (13)	6	Andheri More Sonia Gandhi Camp Saidulajab Village Saheed Camp Jhuggi Lal Bahadur Camp Kishangarh Mangla Puri
Najafgarh	Bijwasan (7)	4	Issapur Khera Bharthal Dhulsiras Bamnauli
Najafgarh	Mundka (12)	6	Bakkarwala Swaran Park Mundka Rly Colony New Chander Vihar Shivram Park Rajdhani Park
West	Tilak Nagar (14)	6	Mukherjee Park Krishna Puri Kesho Pur Major Bhupinder Singh Nagar Krishna Park Mahavir Nagar
West	Uttam Nagar (16)	6	Matiala Nawada Sewak Park Mahendra Park J J Colony Hastal

Table 2.3. South India sample design

Rural component



S.No	Name of Village
1	Pudupattu
2	Velur
3	Chunampet
4	Manappakkam
5	Villipakkam
6	Kavanur
7	Illeedu
8	Pudupettai
9	Othivilakkam
10	V. K. Agaram
11	Thangal
12	Asappur
13	Alathur
14	Vanniya Nallur
15	Arasur
16	Nugumbal
17	Kolatha Nallur
18	Puthiran Kottai
19	Agaram
20	Kayanallur
21	Kalpattu
22	Andarkuppam
23	Mampattu
24	Essur
25	Thuraiyur

Urban component

Chennai Municipal Corporation

↓

15 Zones

↓

200 Corporation wards

↓

40 wards (randomly selected)

↓

10 Census Enumeration Blocks (CEB) per ward -randomly selected (total 400 CEBs)

↓

50 Households per CEB –Systematically selected (total 20000 households)

↓

1 Participant per household- WHO KISH method (total 20000 study participants)

Table 2.4. Pakistan sample design.

Districts in Punjab

1. Attock	19. Lodhran
2. Bahawalnagar	20. Mandi Baha ud din
3. Bahawalpur	21. Mianwali
4. Bhakkar	22. Multan
5. Chakwal	23. Muzaffargarh
6. Chiniot	24. Nankana Sahib
7. D.G.Khan	25. Narowal
8. Faisalabad	26. Okara
9. Gujranwala	27. Pakpattan
10. Gujrat	28. Rahim Yar Khan
11. Hafizabad	29. Rajanpur
12. Jhang	30. Rawalpindi
13. Jhelum	31. Sahiwal
14. Kasur	32. Sargodha
15. Khanewal	33. Sheikhupura
16. Khushab	34. Sialkot
17. Lahore	35. Toba Tek Singh
18. Layyah	36. Vehari

Sites visit plan

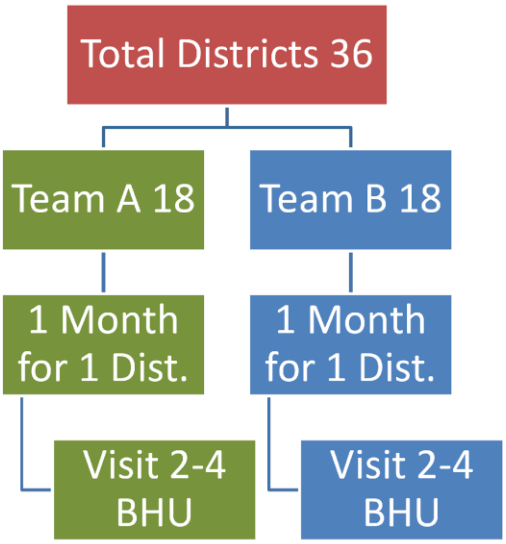


Table 2.5. Sri Lanka sample design

Province and district	Population ('000)	Proportion of SL total	No of Clusters allocated	Distribution of clusters		
				Urban	Rural	Estate
Western	5851	28.7	28			
Colombo	2324		12	9	3	
Gampaha	2305		10	2	8	
Kalutara	1222		6	1	5	
Central	2572	12.6	12			
Kandy	1375		6	1	5	
Matale	485		2	0	2	
Nuwara-eliya	712		4		2	2
Southern	2477	12.1	12			
Galle	1063		6	1	5	
Matara	814		4	1	3	
Hambantota	600		2		2	
Northern	1061	5.4	6			
Jaffna	584		2		2	
Mannar	100		1		1	
Vavuniya	172		1		1	
Mullaitivu	92		1		1	
Kilinochchi	114		1		1	
Eastern	1556	7.6	8			
Batticaloa	527		3	1	2	
Ampara	649		3	1	2	
Trincomalee	380		2		2	
North-western	2381	11.7	12			
Kurunegala	1618		8		8	
Puttalam	762		4		4	
North-central	1267	6.2	6			
Anuradhapura	861		4		4	
Polonnaruwa	406		2		2	
Uva	1266	6.2	6			
Badulla	815		4		3	1
Monaragala	451		2		2	
Sabaragamuwa	1929	9.5	10			
Ratnapura	1088		6	1	5	
Kegalle	841		4		4	
				18	79	3