# BMJ Open

# Cohort profile: the Baependi Heart Study—a family-based, highly admixed cohort study in a rural Brazilian town

Kieren J Egan,[1] Malcolm von Schantz,[1,2,3] André B Negrão,[2] Hadassa C Santos,[2] Andréa R V R Horimoto,[2] Nubia E Duarte,[2] Guilherme C Gonçalves,[2] Júlia M P Soler,[4] Mariza de Andrade,[5] Geraldo Lorenzi-Filho,[6] Homero Vallada,[3] Tâmara P Taporoski,[3] Mario Pedrazzoli,[7] Ana P Azambuja,[8] Camila M de Oliveira,[2,9] Rafael O Alvim,[9] José E Krieger,[2] Alexandre C Pereira[2]

For numbered affiliations see end of article.

**Correspondence to**
Dr Malcolm von Schantz;
m.von.schantz@surrey.ac.uk

## ABSTRACT

**Purpose:** Cardiovascular disease (CVD) is a major challenge to global health. The same epidemiological transition scenario is replayed as countries develop, but with variations based on environment, culture and ethnic mixture. The Baependi Heart Study was set up in 2005 to develop a longitudinal family-based cohort study that reflects on some of the genetic and lifestyle-related peculiarities of the Brazilian populations, in order to evaluate genetic and environmental influences on CVD risk factor traits.

**Participants:** Probands were recruited in Baependi, a small rural town in the state of Minas Gerais, Brazil, following by first-degree and then increasingly more distant relatives. The first follow-up wave took place in 2010, and the second in 2016. At baseline, the study evaluated 1691 individuals across 95 families. Cross-sectional data have been collected for 2239 participants.

**Findings to date:** Environmental and lifestyle factors and measures relevant to cardiovascular health have been reported. Having expanded beyond cardiovascular health outcomes, the phenotype datasets now include genetics, biochemistry, anthropometry, mental health, sleep and circadian rhythms. Many of these have yielded heritability estimates, and a shared genetic background of anxiety and depression has recently been published. In spite of universal access to electricity, the population has been found to be strongly shifted towards morningness compared with metropolitan areas.

**Future plans:** A new follow-up, marking 10 years of the study, is ongoing in 2016, in which data are collected as in 2010 (with the exception of the neuropsychiatric protocol). In addition to this, a novel questionnaire package collecting information about intelligence, personality and spirituality is being planned. The data set on circadian rhythms and sleep will be amended through additional questionnaires, actimetry, home sleep EEG recording and dim light melatonin onset (DLMO) analysis. Finally, the anthropometric measures will be expanded by adding three-dimensional facial photography, voice recording and anatomical brain MRI.

## Strengths and limitations of this study

- High degree of admixture and traditional lifestyle, both uniquely typical of rural Brazil.
- Multiple phenotypes of different categories coupled with genomic and metabolomic data.
- Family-based study design adds heritability information to all variable collected.
- Limitations created by the legal preclusion of payment to participants are circumvented by high compliance within the community, but limits the possibility for complex and/or intrusive study protocols.
- Significant levels of illiteracy complicates the collection of some datasets, and the use of scribes may cause reporting bias.

## INTRODUCTION

Owing to multiple and different waves of immigration, the Brazilian population comprises of a wide variety of ethnic backgrounds. Over successive generations, a considerable amount of admixture has arisen within the population, primarily involving the first groups to arrive (native Americans, Europeans, and Africans).[1–3] In the 2010 census, 43.1% of the population defined themselves as mixed race, 47.7% as white, 7.6% as black, 1.1% as 'yellow' (of Asian descent), and 0.4% as indigenous.[4]

Similar to many developed countries, cardiovascular disease (CVD) is a major non-communicable disease in Brazil.[5] CVD is the leading cause of disability and death, and current statistics suggest that it accounts for 31.2% of deaths worldwide[6] and 31.5% in Brazil.[7] Despite a number of healthcare breakthroughs within the last decade, including more efficient diagnostic and therapeutic procedures, CVD poses a considerable challenge. A number of modifiable

environmental risk factors have been identified (eg, smoking, diet and physical exercise); however, improved awareness does not necessarily equate to behavioural change.[8] Therefore, an improved understanding of the genetics behind CVD, and its interaction with environmental factors, would both improve our understanding of CVD and help prioritise public health resources.

The aetiology of CVD is complex, and involves metabolic, neuroendocrine and genetic interactions.[9–12] Despite a number of lines of evidence suggesting genetic influences, there has been an inconsistency of findings across studies and it is still unknown to what extent the proposed genetic effects are occurring through major loci or multiple distinct loci acting in concert.

Brazil lends itself well to study the environmental contributions of CVD risk profiles. The rapid population growth and industrialisation of Brazil in recent years has established some of the most populated urbanised areas in the world (84% of the population live in urban environments),[4] creating an environmental and lifestyle gap between the metropolitan populations and the still sizable rural ones. Similar to other developed countries, studies have suggested a relatively high prevalence of poor diet,[7] sedentary lifestyle and insufficient exercise.[13] Accordingly, 52.5% of Brazilians are overweight, and a subset of these (17.9% of the general population) are classified as obese.[14]

A substantial element of admixture is both an emblematic feature of the Brazilian population and a facilitating factor in the search for genetic associations. Admixed populations have long been recognised as offering a shortcut towards identifying genomic regions associated with disease or other inheritable traits,[15] an approach known as 'mapping by admixture disequilibrium'.[16] The original aim of this project was to develop a longitudinal family-based cohort study that reflects on some of the genetic and lifestyle-related peculiarities of the Brazilian population in order to help evaluate genetic and environmental influences on cardiovascular risk factor traits. The overall goal was to quantify and characterise the interindividual variation in common cardiovascular risk factors, and disentangle its genetic and environmental components. By collecting and analysing a uniquely wide range of phenotypic and phenomic information, the study aims to serve as a generator of hypotheses for future investigation.

## Cohort description

The study focuses on the population of Baependi, a town in a rural area (752 km$^2$, 18 307 inhabitants at the 2010 census) located in the state of Minas Gerais, Brazil (21.95°S, 44.88°W). It is a traditional community, with a cohesive culture, a high degree of admixture and very limited migration (table 1). Median monthly income at the census date was BRL 510, and 28.1% of inhabitants aged 60 or above and 7.7% of those aged between 24 and 59 years old were illiterate.[4] No individual living in the town was born in another country, and 99.0% of the

population was classified as being born in the South-East region of Brazil.[17]

The initial data collection phase of the Baependi Heart Study took place between December 2005 and January 2006. Prior to recruitment, the project was advertised through provincial, religious and municipal authorities, in local television, newspaper and radio messages, through physicians, and by phone calls. For physical examination, a clinic was established in the centre of Baependi. Currently, a general practitioner, a physician, funded through the project is operating a clinic in these premises, where health concerns identified through the study are followed up.

**Table 1** Basic description of the study population and comparison with the 2010 Census data

| Totals | Baependi study N (%) | Census N (%) |
|---|---|---|
| Sex (data representing 2239/2239 individuals) | | |
| Male | 977 (43.6) | 9175 (50) |
| Female | 1262 (56.4) | 9132 (50) |
| Age over 18 (data representing 2239/2239 individuals | | |
| 18 to 19 | 87 (3.9) | 623 (3.4) |
| 20 to 24 | 204 (9.1) | 1460 (8) |
| 25 to 29 | 219 (9.8) | 1454 (7.9) |
| 30 to 39 | 395 (17.6) | 2549 (13.9) |
| 40 to 49 | 419 (18.7) | 2646 (14.5) |
| 50 to 59 | 412 (18.4) | 2158 (11.8) |
| 60 and over | 503 (22.5) | 2606 (14.2) |
| Place of Birth (data representing 1893/2239 individuals) | | |
| South-East Brazil | 1826 (99.3) | 18121 (99) |
| Brazil | 1893 (100) | 18307 (100) |
| Education (data representing 2235/2239 individuals) | | |
| **No schooling** | | |
| Illiterate | 115 (5.1) | - |
| literate | 106 (4.7) | - |
| **Schooling (years)** | | |
| 1 to 4 | 712 (31.9) | - |
| 5 to 8 (end of Primary schooling) | 412 (18.4) | - |
| 9 to 11 (end of secondary schooling) | 583 (26.1) | - |
| >11 | 156 (7.0) | - |
| Finished University | 151 (6.8) | - |
| Employment status (data representing 2229/2239 individuals | | |
| Employed | 1198 (53.7) | * |
| Seeking employment | 76 (3.4) | * |
| Economically inactive | 955 (42.8) | * |
| Ethnicity (data representing 2232/2239 individuals) | | |
| White | 1711 (76.7) | 13 669 (74.7) |
| Black | 124 (5.6) | 1032 (5.6) |
| Mixed | 384 (17.2) | 3554 (19.4) |
| Asian | 12 (0.54) | 46 (0.3) |
| Indigenous | 1 (0.04) | 6 (0.03) |

Table shows baseline data compared with census data.
*Numbers are not comparable due to the exclusion of individuals below 18 years of age from the study.

Probands were identified from the community in 11 census districts (from a possible 12). Residential addresses within each district were then randomly selected (by randomly selecting first a street, and second a household on that street). Once a proband in this household, 18 years or older, was enrolled, all his/her first-degree (eg, parents, siblings and offspring), second-degree (eg, half-siblings, grandparents/grandchildren, aunts/uncles, nieces/nephews and double cousins), and third-degree (eg, first cousins, great-uncles/great-aunts, and great-nephews/great-nieces) relatives and his/her respective spouse's relatives, who were at least 18 years old, were invited to participate. After the proband's first contact, first-degree relatives were invited to participate by phone; these included all living relatives in the municipality of Baependi (the urban zone, where 72% of the population lives[4] and the geographically larger rural zone) and beyond. Ninety-five families were selected, and with 1691 individuals who have participated in some or all of the measurements of the study; they comprise more than 10% of the local population (see figure 1 for overview of study population).

The mean pedigree size was 24.2±31.8, and most of them (63%) encompassed three or four generations. There were 640 sibships with a mean size of 2.4±1.9 and the following numbers of main pairs of relatives: 3138 parent/offspring, 2253 sibling, 2590 grandparent/grandchild, 4418 avuncular, 40 half-siblings and 3743 cousins.

Medical history was collected from all participants at baseline. Described in the following are outcomes (incidence information derived from longitudinal data collection) and exposures (cross-sectional data collected at baseline or at subsequent study visits). The first 5-year follow-up took place in 2010, and alongside this, specific data collections have been added. The 2010 follow-up involved both renewed DNA sampling and measures of major cardiovascular risk factors. Cardiovascular events (end points such as myocardial infarction, heart failure, coronary insufficiency) and procedures (such as hospitalisations, surgery and the need for percutaneous coronary intervention) are recorded regularly; individuals are followed up annually by telephone contact and every 5 years during office visits. The attrition rate (based on deaths and follow-up refusals) was 23%. In the 2010 follow-up, 548 individuals were added (figure 2).

A new follow-up, marking 10 years of the study, is ongoing, in which data are collected as in 2010 (with the exception of the neuropsychiatric protocol). Taken together, baseline measurements represent 2239 individuals. For follow-up percentages of individual outcomes, see table 2.
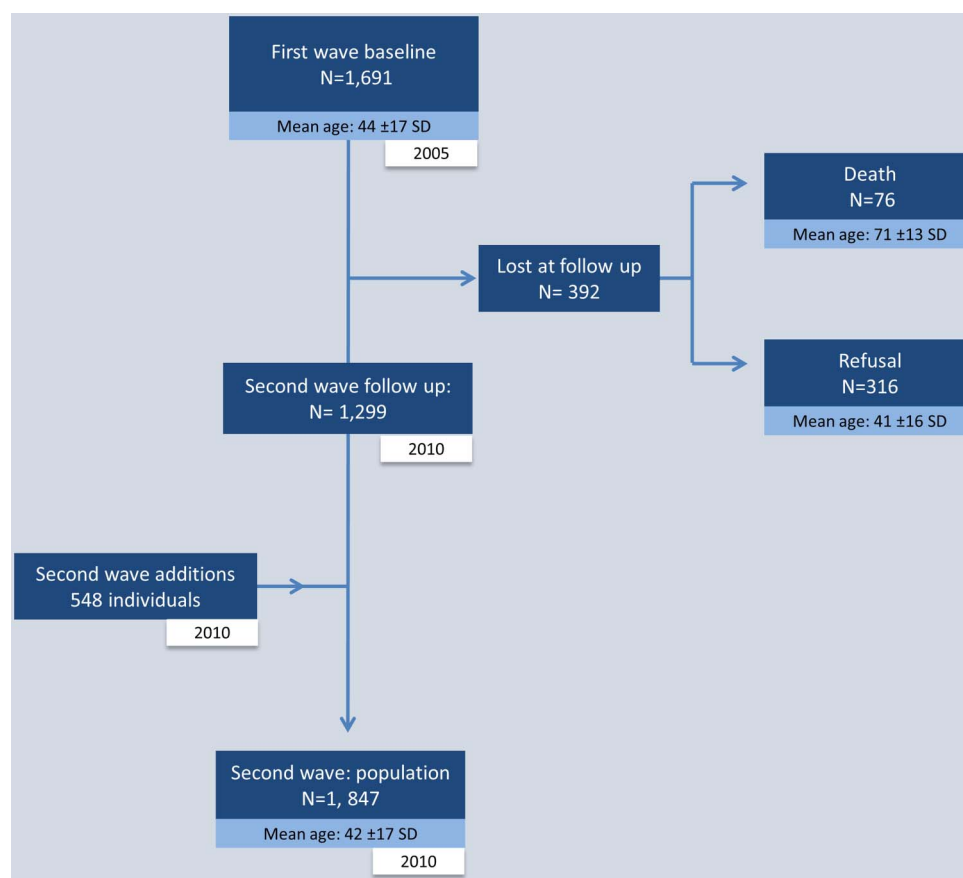


**Figure 1** Flow diagram of study population at baseline and follow-up.
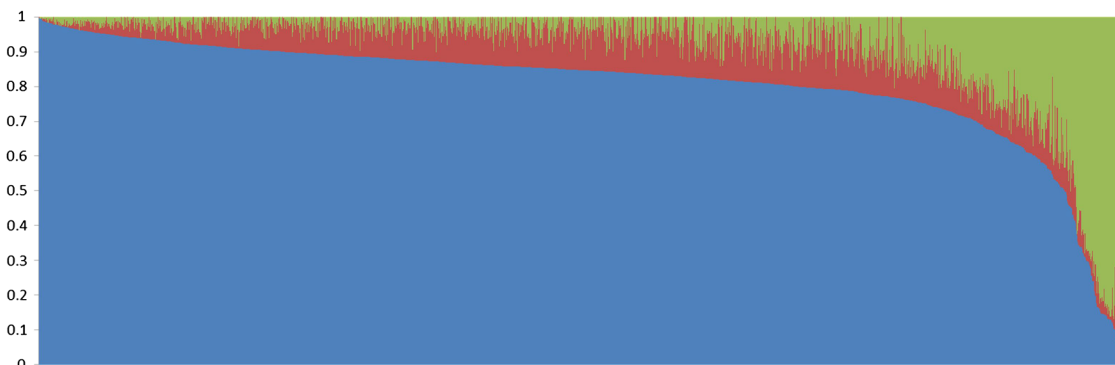
**Figure 2** Admixture proportions for the Baependi cohort inferred by the ADMIXTURE software. The X-axis represents individuals from Baependi sorted according to their ancestries. Each individual is represented by a vertical column of colour-coded admixture proportions of the ancestral populations (from HapMap), CEU (Utah residents (CEPH)) and Toscani in Italia (TSI) (European, blue), Human Genome DIversity Project (HGDP), Pima and Maya (Native American, red) and Yoruba in Ibadan, Nigeria (YRI), Luhya in Webuye, Kenya (LWK) and Americans of African ancestry in SW USA (ASW) (African, green).

A general questionnaire was initially administered to each participant by a trained technician to cover basic attributes such as family relationships and demographic characteristics. For participants who were illiterate or visually impaired, questionnaires were read in their entirety by the researcher acting as a scribe. Where an individual's capacity was limited to complete a specific question, assistance was provided as needed by either the researcher or a caregiver.

The investigators also used the WHO MONICA epidemiological instrument[18] which had previously been successful within other epidemiological projects. While the core data about cardiovascular risk factor prevalence and heritability were collected at baseline as well as at follow-up, and have been described in detail previously,[19] additional parameters have subsequently been recorded pertaining to biochemistry, sleep and mental health. For a full list of study variables measured see table 2. The following types of information were collected (for details of variables and distribution estimates, see table 3): general lifestyle and health, cardiovascular health, biochemistry, anthropometry, mental health and sleep and circadian rhythms.

In addition, leucocyte DNA samples have been collected and used to probe Affymetrix 6.0 arrays, and data imputed using the 1000 Genomes Cosmopolitan Panel. Analysis of genomic ancestry was conducted using the Admixture software.[20]

Each participant provided informed written consent before participation.

## Findings to date

The study has begun to make significant contributions to the understanding of a variety of conditions including the heritability of risk factors for cardiovascular events. Many of these have yielded heritability estimates, summarised in table 4. The initial publication[19] determined the extent to which unmeasured genetic factors and measured environmental and lifestyle factors contributed to variation in a large panel of cardiovascular-related traits. Subsequent studies identified that age-at-onset is a useful trait for gene mapping of common complex diseases[21] and reported that heterogeneity in trait variances needs to be accounted for in the design and analyses of trait orientated gene finding studies.[22] Lifestyle factors, including physical activity[23] and smoking[24] have also been a focus of the study.

The study has been extended to also include aspects of mental health, another major non-communicable disease complex forming a large burden to society. The most recent report emanating from the study showed that anxiety and depression, measured as continuous variables within the general population rather than as clinical diagnoses, shared a significant genetic background.[25] Skin properties of the study participants have also been collected, and a report of significant associations between stratum corneum moisture and sex, age, high sun exposure, and use of sunscreen has been published.[26] Sleep and circadian rhythms, having important links with cardiovascular health,[27 28] are a focus of the project, including a general screen for sleep apnoea which was diagnosed in 18.6% of the population and had similar heritability estimates compared with other urban populations studies (See table 4).[29] A study of diurnal preference (chronotype) which indicated that, in spite of universal access to electricity, the Baependi population was strongly shifted towards morningness, particularly in the rural zone.[17]

In parallel with these primary interests have been developments in statistical methodologies due to the family-based structure of the study.[30 31] For example, the use of genome-wide association study (GWAS) techniques in family-based studies is susceptible to confounders. Therefore researchers used simulations from two datasets including the Baependi study to derive coefficients which incorporate the relatedness of individuals within family-based designs.[31] These estimates suggest

**Table 2** Follow-up rate across core variables

| Outcome | Scale or measures | Number at baseline | Follow-up rate |
|---|---|---|---|
| Anthropometry | | | |
| Weight (kg) | | 1654 | 72% |
| Height (m) | | 1654 | 72% |
| BMI (kg/m$^2$) | | 1654 | 72% |
| Waist circumference (cm) | | 1689 | 78% |
| Hip circumference (cm) | | 1725 | NA* |
| Neck circumference (cm) | | 1739 | NA* |
| General lifestyle | | | |
| Smokers[24] | Initiation, persistence, quantity, age at onset of regular cigarette use | 1690 | 77% |
| Sedentarism | IPAQ (International Physical Activity Questionnaire, short version)[35] | 1688 | 75% |
| Physical activity | IPAQ (short version)[35] | 1690 | 77% |
| Nutritional assessment | Food frequency questionnaire[36] | 1819 | NA* |
| Alcohol consumption | Self-report | 1682 | 77% |
| Socioeconomic profile | Self-report | 1690 | 77% |
| Education level | Self-report | 1690 | 77% |
| Cardiovascular health | | | |
| 24-hour ambulatory BP monitoring (mm Hg) | | 1351 | NA* |
| Carotid–femoral pulse wave velocity (m/s) | | 1675 | NA* |
| Ankle–brachial index (ratio) | | 1631 | NA* |
| Carotid intima-media thickness (mm) | | 1735 | NA* |
| Heart rate variability in 2 positions (supine and standing, ms/un and ms$^2$) | | 1735 | NA* |
| Systolic BP (mm Hg) | | 1686 | 72% |
| Diastolic BP (mm Hg) | | 1684 | 72% |
| 24-hour (Holter monitoring) | | 1059 | NA* |
| Vascular age (years) | | 1651 | NA* |
| Biochemical measurements | | | |
| Fasting glucose (mg/dL) | | 1626 | 74% |
| Triglycerides (mg/dL) | | 1661 | 74% |
| HDL-c (mg/dL) | | 1666 | 74% |
| LDL-c (mg/dL) | | 1653 | 73% |
| Total cholesterol (mg/dL) | | 1662 | 74% |
| Fasting metabolome (mg/dL) | | 1485 | NA* |
| HbA1c (%) | | 1772 | NA* |
| Skin phenotypes | | | |
| Dermatology | Visual imaging[37], optical coherence tomography (algorithms), Corneometer (scale application), transepidermal water loss (scale application), Cutometer (scale application) questionnaires[38] | 1587 | NA* |
| Mental health | | | |
| Symptoms for major psychiatric disorders | CIS-R[39] | 1701 | NA* |
| Cognitive and executive function | Words or figure memory test; semantic (animals) and phonemic (letter F) verbal fluency tests; Trail Making Test | 1701 | NA* |
| Executive function | Random Number Generation, plus/minus | 1301 | NA |
| Anxiety | Hospital Anxiety and Depression Scale[40] | 1375 | NA* |
| Depression | Hospital Anxiety and Depression Scale[40] | 1375 | NA* |
| Sleep and circadian rhythms | | | |
| Sleep | Pittsburgh Sleep Quality Index,[41] Epworth Sleepiness Questionnaire,[42] Berlin Questionnaire[43] | 1758 | NA* |
| Diurnal preference | Morningness–eveningness questionnaire[44] | 1327 | NA* |

*Follow-up was not possible as baseline information was only collected in 2010.
BMI, body mass index; BP, blood pressure; CIS-R, Revised Clinical Interview Schedule; HbA1c, glycated haemoglobin; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; NA, not applicable.

**Table 3** Descriptive statistics of studied phenotypes, distributed by gender (cross-sectional type data, parentheses show SD)

|  | Total N=2239 | Men N=977 | Women N=1262 |
|---|---|---|---|
| Age (years) | 42.5 (17.2) | 42.8 (17.8) | 42.4 (16.8) |
| Waist circumference (cm) | 95.8 (11.3) | 94.0 (9.9) | 97.3 (12.1) |
| Fasting glucose (mg/dL) | 92.7 (26.7) | 92.3 (23.6) | 93.0 (28.8) |
| Triglycerides (mg/dL) | 132.0 (78.6) | 135.0 (90.6) | 129.7 (67.7) |
| HDL-c (mg/dL) | 53.8 (15.2) | 51.4 (14.9) | 55.6 (15.2) |
| Total cholesterol (mg/dL) | 184.7 (47.5) | 180.1 (47.9) | 188.2 (46.8) |
| LDL-c (mg/dL) | 104.9 (44) | 101.9 (44.1) | 107.1 (43.8) |
| Systolic BP (mm Hg) | 125.9 (18.6) | 129.8 (17.8) | 122.9 (18.6) |
| Diastolic BP (mm Hg) | 77.8 (11.2) | 78.7 (11.6) | 77.1 (10.9) |
| BMI (kg/m$^2$) | 24.6 (4.9) | 23.7 (4.2) | 25.3 (5.3) |
| High fasting glucose (>110 mg/dL) | 14.5% | 12.8% | 15.8% |
| High BP (mean systolic BP of >140 mm Hg and/or diastolic BP of >90 mm Hg or the use of antihypertensive medications) | 35.2% | 34.1% | 36.1% |
| Truncal obesity | 29.9% | 8.7% | 46.4% |
| High triglycerides | 27.8% | 28.2% | 27.5% |
| Low HDL-c (<40 mg/dL in men and <50 mg/dL in women) | 31.7% | 22.1% | 39.1% |
| High LDL-c (>130 mg/dL) | 26.3% | 23.9% | 28.1% |
| BMI>25 kg/m$^2$ | 39.7% | 32.2% | 45.5% |
| BMI>30 kg/m$^2$ | 13.3% | 8.2% | 17.3% |
| Smokers | 15.4% | 19.0% | 12.7% |

BMI, body mass index; BP, blood pressure; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol.

**Table 4** Estimates from heritability studies and comparisons with previously published estimates

| Variable | Adjusted heritability Percentage (SEM) | Range of heritability values in published literature | Reference |
|---|---|---|---|
| Cardiovascular disease |  |  |  |
| Waist circumference (cm) | 40.1 (0.049) | 5.8–54 | [19] |
| Fasting glucose (mg/dL) | 32.3 (0.047) | 18–39 | [19] |
| Triglycerides (mg/dL) | 28.4 (0.048) | 19–56 | [19] |
| HDL-cholesterol (mg/dL) | 32.3 (0.046) | 29.9–62 | [19] |
| Systolic blood pressure (mm Hg) | 20.6 (0.042) | 16–48 | [19] |
| Diastolic blood pressure (mm Hg) | 21.0 (0.044) | 21–34 | [19] |
| Total cholesterol (mg/dL) | 30.4 (0.049) | 32.8–32.8 | [19] |
| LDL-cholesterol (mg/dL) | 27.2 (0.050) | 33–40.1 | [19] |
| BMI (kg/m$^2$) | 51.0 (0.048) | 21.6–51.0 | [19] |
| Physical activity |  |  |  |
| Weekly physical activity | 11.0 (0.04) | 6–83 | [23] |
| Sedentarism | 12.0 (0.07) |  |  |
| Physical activity+non-sedentary | 35.0 (0.11) |  |  |
| Cigarette use |  |  |  |
| Smoking initiation | 69.6 (0.123) | 11–75 | [24] |
| Smoking persistence | 31.9 (0.145) | 52–71 | [24] |
| Smoking quantity | 25.6 (0.098) | 51–57 | [24] |
| Obstructive sleep apnoea |  |  |  |
| Obstructive Sleep apnoea | 25.0 | 21–37 | [29] |
| Diurnal preference |  |  |  |
| Diurnal preference | 38.0 (0.09) | 21–52 | [17] |
| Depression and anxiety |  |  |  |
| Anxiety | 32.0 (0.06) | 25–60 | [25] |
| Depression | 30.0 (0.60) | 17–78 | [25] |

BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

that family structure is important for the estimation of global individual ancestry for extended pedigrees, but not for siblings. A second key statistical output from the cohort concerns the complexity of multiple CVD phenotypes, which makes disease diagnosis and genetic dissection difficult. Using the cohort data the groups were able to propose empirically guided statistical methodologies which account for all categorical phenotypes, allowing informative comparison among individuals.[30]

### Strengths and limitations

Over the last decade, the Baependi cohort study has begun to fulfil its potential of making significant contributions to our understanding of the heritability across a variety of different health conditions. A number of study attributes make this study population unique, including the substantial admixture, the conservative and cohesive lifestyle patterns, and the increasingly rich tapestry of both cardiovascular and other variables extracted at baseline and follow-up.

The researchers did not set out with a specific target population number and in contrast to the design of some longitudinal study designs, there was no specific intention to provide a representative sample of the local population. However, the study sample was obtained using randomised procedures, whereby the researchers attempted to achieve equal representation of all socioeconomic groups. Equally, the selection was not specific to any specific phenotype of interest, and therefore our sample is relatively free from disease-specific bias. Further, the multigenerational nature of the study means that it exceeds the conventional nuclear family unit.

The legal preclusion in Brazil of offering payments to study volunteers is a general limitation, especially for more complex and intrusive study protocols. Nonetheless, thanks to the engagement and support of the community, and the provision of a general practitioner in the field station, the retention rate of the study remains very high.[32]

Although the wide variety of outcomes recorded is a key strength of the study, the extensive use of scribes and face-to-face interviews may have exacerbated reporting bias (eg, for diurnal preference, smoking, alcohol). To control for outcome variability, wherever possible, multiple measurements were collected (eg, systolic and diastolic blood pressure[19]). However, some analyses were based on single samples (eg, tryglycerides[19]), whereas repeated measurements are likely to have given more reliable estimates of individual range.

The family-based study design is noteworthy because it adds information (family structure) to any variable that is statistically estimated as opposed to cohorts with independent participants; in practice, heritability estimates can be generated from any desired variable. Although further follow-up of individuals is ongoing (and further follow-ups are anticipated), the 5 years follow-up is insufficient for long-term trajectory of individuals. Another consideration is that the local population has a relatively low level of education and income. These covariates have been considered carefully, particularly where lower socioeconomic status is known to raise the risk for specific health conditions such as anxiety and depression.[33][34]

### COLLABORATION

The database of the project is under continuous development as new datasets are being added. Data will be made available in public repositories as and when required and appropriate, as publications based on datasets are being accepted. The project already involves multiple collaborations, and the investigators welcome new angles of analyses on existing data sets and proposals for new ones. Proposals can be forwarded to ACP (email: alexandre.pereira@incor.usp.br) for discussion at the steering committee meeting.

**Author affiliations**
[1]Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK
[2]Laboratory of Genetics and Molecular Cardiology, Heart Institute (Incor), University of São Paulo Medical School, São Paulo, Brazil
[3]Institute of Psychiatry, University of São Paulo Medical School, São Paulo, Brazil
[4]Department of Statistics, University of São Paulo, São Paulo, Brazil
[5]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA
[6]Sleep Laboratory, Pulmonary Division, Heart Institute (Incor), University of São Paulo Medical School, São Paulo, Brazil
[7]School of Arts, Science, and Humanities, University of São Paulo, São Paulo, Brazil
[8]Natura Innovation and Product Technology Ltd., Cajamar, SP, Brazil
[9]Department of Physiology, Federal University of Juiz de Fora, Juiz de Fora, Brazil

# REFERENCES

1. Kehdy FSG, Gouveia MH, Machado M, *et al*. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci USA* 2015;112:8696–701.
2. Ribeiro D. *O povo brasileiro: a formação e o sentido do Brasil*. Sao Paulo: Companhia das Letras, 1995.
3. Carvalho-Silva DR, Santos FR, Rocha J, *et al*. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 2001;68:281–6.
4. Instituto Brasileiro de Geografia e Estastı́stica, Census data (http://censo2010.ibge.gov.br/). 2010.
5. Schmidt MI, Duncan BB, Azevedo e Silva G, *et al*. Health in Brazil 4 chronic non-communicable diseases in Brazil: burden and current challenges. *Lancet* 2011;377:1949–61.
6. World Health Organization. Published Online First: 2015. http://www.who.int/mediacentre/factsheets/fs317/en (accessed 24 Aug 2015).
7. Fisberg RM, Mirimoto JM, Slater B, *et al*. Dietary quality and associated factors among adults living in the state of São Paulo, Brazil. *J Am Diet Assoc* 2006;106:2067–72.
8. Alzaman N, Wartak SA, Friderici J, *et al*. Effect of patients' awareness of CVD risk factors on health-related behaviors. *South Med J* 2013;106:606–9.
9. Steptoe A, Wardle J, Marmot M. Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. *Proc Natl Acad Sci USA* 2005;102:6508–12.
10. Mottillo S, Filion KB, Genest J, *et al*. The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. *J Am Coll Cardiol* 2010;56:1113–32.
11. Kathiresan S, Srivastava D. Genetics of human cardiovascular disease. *Cell* 2012;148:1242–57.
12. Goldin LR, Camp NJ, Keen KJ, *et al*. Analysis of metabolic syndrome phenotypes in Framingham Heart Study families from genetic analysis workshop 13. *Genet Epidemiol* 2003;25:S78–89.
13. Dumith SC. Physical activity in Brazil: a systematic review atividade física no Brasil: uma revisão sistemática. *Cad Saude Publica* 2009;25:415–26.
14. Brazilian Ministry of Health Report [Vigitel Brasil 2014]: Vigilância de Fatores de Risco e Proteção Para Doenças Crônicas por Inquérito Telefônico. Published Online First. 2014. http://www.abeso.org.br/uploads/downloads/72/553a243c4b9f3.pdf (accessed on 9 Oct 2015).
15. Rife DC. Populations of hybrid origin as source material for the detection of linkage. *Am J Hum Genet* 1954;6:26–33.
16. Stephens JC, Briscoe D, Brien SJ. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am J Hum Genet* 1994;55:809–24.
17. von Schantz M, Taporoski T, Horimoto A, *et al*. Distribution and heritability of diurnal preference (chronotype) in a rural Brazilian family-based cohort, the Baependi study. *Sci Rep* 2015;5:9214.
18. The World Health Organization MONICA project (monitoring trends and determinants in cardiovascular disease): a major collaboration. WHO MONICA Project Principal Investigators. *J Clin Epidemiol* 1988;41:105–14.
19. de Oliveira CM, Pereira AC, de Andrade M, *et al* Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC Med Genet* 2008;9:32.
20. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655–64.
21. Ruiz Giolo S, Pereira AC, de Andrade M, *et al*. Genetic analysis of age-at-onset for cardiovascular risk factors in a Brazilian family study. *Hum Hered* 2009;68:131–8.
22. Giolo SR, Pereira AC, de Andrade M, *et al*. Evaluating gene by sex and age interactions on cardiovascular risk factors in Brazilian families. *BMC Med Genet* 2010;11:132.
23. Horimoto ARVR, Giolo SR, Oliveira CM, *et al*. Heritability of physical activity traits in Brazilian families: the Baependi Heart Study. *BMC Med Genet* 2011;12:155.
24. Horimoto ARVR, Oliveira CM, Giolo SR, *et al*. Genetic analyses of smoking initiation, persistence, quantity, and age-at-onset of regular cigarette use in Brazilian families: the Baependi Heart Study. *BMC Med Genet* 2012;13:9.
25. Taporoski TP, Negrão AB, Horimoto ARVR, *et al*. Shared genetic factors of anxiety and depression symptoms in a Brazilian family-based Cohort, the Baependi Heart Study. *PLoS ONE* 2015;10:e0144255.
26. de Farias Pires T, Azambuja AP, Horimoto AR, *et al*. A population-based study of the stratum corneum moisture. 2016;9:79–87.
27. Rangaraj VR, Knutson KL. Association between sleep deficiency and cardiometabolic disease: implications for health disparities. *Sleep Med* 2016;18:19–35.
28. Portaluppi F, Tiseo R, Smolensky MH, *et al*. Circadian rhythms and cardiovascular health. *Sleep Med Rev* 2012;16:151–66.
29. de Paula LK, Alvim RO, Pedrosa RP, *et al*. Heritability of obstructive sleep apnea in a rural population. *Chest* 2016;149:92–7.
30. Fragoso TM, Giolo SR, Pereira AC, *et al*. Using item response theory to model multiple phenotypes and their joint heritability in family data. *Genet Epidemiol* 2014;38:152–61.
31. de Andrade M, Ray D, Pereira AC, *et al*. Global individual ancestry using principal components for family data. *Hum Hered* 2015;80:1–11.
32. Booker CL, Harding S, Benzeval M. A systematic review of the effect of retention methods in population-based cohort studies. *BMC Public Health* 2011;11:249.
33. Melchior M, Chastang JF, Walburg V, *et al*. Family income and youths' symptoms of depression and anxiety: a longitudinal study of the French Gazel youth cohort. *Depress Anxiety* 2010;27:1095–103.
34. Lorant V, Croux C, Weich S, *et al*. Depression and socio-economic risk factors: 7-year longitudinal population study. *Br J Psychiatry* 2008;190:293–8. .
35. Craig CL, Marshall AL, Sjöström M, *et al* International physical activity questionnaire: 12-country reliability and validity. 2003;35:1381–95.
36. Thompson FE, Byers T. Dietary assessment resource manual. *J Nutr* 1994;124:2245S–317S.
37. Fitzpatrick TB. The validity and practicality of sun-reactive skin types 1 through VI. *Arch Dermatol* 1988;124:869–71.
38. Rawlings A V. Cellulite and its treatment. *Int J Cosmet Sci* 2006;28:175–90.
39. Lewis G, Pelosi AJ, Araya R, *et al*. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychol Med* 1992;22:465–86.
40. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983;67:361–70.
41. Buysse DJ, Reynolds CF, Monk TH, *et al*. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193–213.
42. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–5.
43. Netzer NC, Stoohs RA, Netzer CM, *et al*. Using the Berlin Questionnaire to identify patients at risk for the sleep. *Ann Intern Med* 1966;198:223–7.
44. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-evenings in human circadian rhythms. *Int J Chronobiol* 1976;4:97–110.