How did the study come about?
Population Registry of Lifestyle Diseases (PROLIFE) is a prospective cohort study aimed at the long-term follow-up of the residents of seven contiguous villages in Kerala, India. There were 161,942 subject enrolments as of 1 July 2002. The study was initiated by Health Action by People (HAP), a non-profit public health research organization.1

Kerala, with a population of >31 million in 2001, has the best health indicators in India.2,3 The ‘Kerala health experience’ has been much discussed in academic circles.4–6 The dramatic declines in fertility and mortality in the state have resulted in a rapid epidemiological transition characterized by high morbidity and mortality from chronic diseases. However, valid epidemiological information on the burden of chronic diseases in the population does not exist. We have to rely on secondary data from hospitals. HAP decided to bridge the data gap by setting up PROLIFE.

What does the study cover?
At the time of set up, PROLIFE was more in the nature of an exploratory study and not based on a single hypothesis. Our first goal was to assemble a cohort to study the epidemiology of chronic diseases in the selected rural community. In its second phase, the study proposes to estimate the burden of risk factors, both conventional and novel, and look for association with major causes of death in Kerala. We also propose to investigate the social determinants of diseases and deaths in our community. As we are collecting detailed information on births and deaths, life course epidemiology would be a future focus of the study.

How is the study set up and who are in the sample?
The study is set up in the Varkala rural development block in Thiruvananthapuram district, Kerala (Figure 1). The Varkala block comprises seven panchayats (local governments) with population ranging from 15,869 to 28,309. These panchayats are served by 105 childcare centres (anganvadis), as part of India’s ‘Integrated Child Development Services’. Each anganvadi caters to approximately 100 children and is managed by a female health worker. These health workers served as enumerators for the study with the consent and support of the local administration. Oral consent was obtained from the head of each household for the initial baseline survey. Written, informed consent was obtained from all respondents in subsequent resurveys. The study was approved by the institutional ethics committee.

All residents living in the Varkala block were eligible to participate in the study. The block has an area of 86.67 km² and the density of population was 1873/km², more than double the state average. The study area was predominantly rural (73% of population) and the per capita annual income was <US$ 1000 per year. Preparatory to the conduct of the survey, we identified all households in the area, totalling 34,190. The health workers visited all households and explained the purpose of the study to the head of each household. Refusals to participate amounted to 2.4%. We enrolled 33,379 households with a population of 161,942 (79,422 men and 82,398 women) in the study. The age structure of the study population closely matched the age composition of Kerala population (Table 1).

A detailed household level survey was carried out between 1 July and 31 December 2001. Among
other things, the survey collected the names, number of residents and number of adults in each household. A food frequency questionnaire was also administered in each household. Subjects aged ≥20 years of age were enrolled into an adult cohort. Separate adult questionnaires were administered to 78,173 subjects (33,978 men and 44,195 women) who were ≥20 years as of 1 January 2002. Baseline data collection was completed on 30 June 2002. Figure 2 presents a flow chart outlining the essentials of recruitment and subsequent activities.

**How are they being followed up?**

Before starting the survey, all health workers (105) and supervisors (7) were trained by physician epidemiologists on survey methods, recognition of common

---

Table 1  Age and gender composition and comparison with Kerala population

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>Males</th>
<th>Females</th>
<th>Total (%)</th>
<th>Kerala (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4</td>
<td>6807</td>
<td>6749</td>
<td>13,556 (8.4)</td>
<td>2,766,154 (8.7)</td>
</tr>
<tr>
<td>5–9</td>
<td>6204</td>
<td>6104</td>
<td>12,308 (7.6)</td>
<td>2,544,181 (8.0)</td>
</tr>
<tr>
<td>10–14</td>
<td>7219</td>
<td>7162</td>
<td>14,381 (8.9)</td>
<td>2,987,200 (9.4)</td>
</tr>
<tr>
<td>15–19</td>
<td>6966</td>
<td>7425</td>
<td>14,391 (8.9)</td>
<td>2,984,506 (9.4)</td>
</tr>
<tr>
<td>20–24</td>
<td>6803</td>
<td>8296</td>
<td>15,099 (9.3)</td>
<td>2,983,990 (9.4)</td>
</tr>
<tr>
<td>25–29</td>
<td>7257</td>
<td>7964</td>
<td>15,221 (9.4)</td>
<td>2,786,195 (8.7)</td>
</tr>
<tr>
<td>30–34</td>
<td>7348</td>
<td>6902</td>
<td>14,250 (8.8)</td>
<td>2,516,463 (7.9)</td>
</tr>
<tr>
<td>35–39</td>
<td>6794</td>
<td>6731</td>
<td>13,525 (8.3)</td>
<td>2,466,354 (7.7)</td>
</tr>
<tr>
<td>40–44</td>
<td>5474</td>
<td>4954</td>
<td>10,428 (6.4)</td>
<td>1,951,284 (6.1)</td>
</tr>
<tr>
<td>45–49</td>
<td>5227</td>
<td>4589</td>
<td>9,816 (6.1)</td>
<td>1,926,144 (6.0)</td>
</tr>
<tr>
<td>50–54</td>
<td>3644</td>
<td>3409</td>
<td>7,053 (4.3)</td>
<td>1,437,520 (4.5)</td>
</tr>
<tr>
<td>55–59</td>
<td>2618</td>
<td>2978</td>
<td>5,596 (3.4)</td>
<td>1,130,244 (3.5)</td>
</tr>
<tr>
<td>60–64</td>
<td>2389</td>
<td>2838</td>
<td>5,227 (3.2)</td>
<td>1,032,136 (3.2)</td>
</tr>
<tr>
<td>65–69</td>
<td>2035</td>
<td>2508</td>
<td>4,543 (2.8)</td>
<td>902,015 (2.8)</td>
</tr>
<tr>
<td>70–74</td>
<td>1307</td>
<td>1678</td>
<td>2,985 (1.8)</td>
<td>613,422 (1.9)</td>
</tr>
<tr>
<td>75–79</td>
<td>822</td>
<td>996</td>
<td>1,818 (1.1)</td>
<td>399,089 (1.2)</td>
</tr>
<tr>
<td>≥80</td>
<td>630</td>
<td>1115</td>
<td>1,745 (1.0)</td>
<td>389,013 (1.2)</td>
</tr>
<tr>
<td>Total</td>
<td>79,544</td>
<td>82,398</td>
<td>161,942</td>
<td>3,184,137 (1.2)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Kerala population taken from census 2001.
diseases and ascertainment of the cause of death, using structured forms. Our health workers visit the allotted households once every 3 months and seek information on birth, death or any other major health event. As the health worker is a local resident, the community alerts her about new births and deaths. This enables the worker to visit the house shortly after the occurrence of any event. Reporting of vital events to the central office is made by telephone on the basis of a pre-designed format. The ongoing repeat survey will capture all subjects who turned 20 years of age since the first survey. An adult life-style questionnaire will be administered to each of them. Migration histories of the subjects are being tracked.

What is the response rate and likely rate of loss?

We administered the adult questionnaire to 78 173 subjects, who constitute 80.4% of eligible members. The shortfall was primarily on account of some subjects being absent from home on three consecutive visits. Other than deaths, migration would be the main reason for loss to follow-up. As of 31 December 2004, 721 subjects moved out of the study area and were lost to follow-up. This constitutes <0.01% of the subjects, over a period of 42 months. As years pass, the study is gaining wider acceptance from the community. The regular presence of physicians and provision of periodic health screening have augmented the acceptability of the study.

What has been measured?

The baseline household survey collected information on demography, quality of housing, asset ownership and details of migration. A food frequency questionnaire sought details relating to the frequency of consumption of 75 common food items and beverages. Information on births, deaths and causes of death was collected for the 5-year period preceding the baseline survey (1996–2001). The adult life style questionnaire captured demographic features, indicators of socio-economic position and history of overseas employment. Detailed history was also sought on the nature and frequency of tobacco and alcohol use, and questions were asked on physical activity, personal medical history and family medical history.

A summary of the questionnaires is presented in Table 2.

In the event of a death, the health worker visited the household and made detailed inquiries into the cause of death. A specially structured questionnaire supported by a symptom list was used to arrive at a probable cause of death. The reported causes of death were reassessed by a trained physician in the coordinating office. The physician assigned an underlying cause of death for each case and allotted it to a broader category based on the Tenth Revision of International Classification of Disease. The sensitivity and specificity of verbal autopsy tools for broad categories are much higher than those for specific causes. We believe that our use of broad categories ensures minimal error in classification of deaths.

We are assessing risk-factor prevalence in selected members of the cohort, based on a cluster sampling approach. Measurement of height, weight, waist circumference, hip circumference, mid-arm circumference, subcutaneous fat over triceps and sitting blood pressure have been completed in 12 692 subjects as of 31 December 2007. Biochemical measurements such as haemoglobin, fasting and 2-h plasma glucose, serum cholesterol, high-density lipid (HDL) cholesterol, triglycerides, uric acid, creatinine and blood urea nitrogen were completed in 8116 subjects. All lipid constituents, uric acid, creatinine and blood urea nitrogen were measured in fasting serum samples. Aliquots of serum were stored at −70°C for future analysis.

A re-survey using the original instruments is currently being carried out. The survey was started in January 2008 and is almost complete at the time of this writing. The survey offers us an opportunity to cross-check on the reliability of annual reporting of births and deaths. We also hope to capture the socio-economic mobility of the households since the original survey.
What has it found?

Analysis of the cause of 3411 deaths that occurred from 1996 to 2001, shows that 49% of deaths in this community are due to cardiovascular diseases (CVDs). This is followed by external causes of mortality (12%) and cancer (9%). Infections contribute to <6% of deaths. Nearly 10% of deaths could not be assigned specific causes. Preliminary analysis of more than 4000 deaths prospectively collected since 2002 indicates excellent level of agreement with the above findings.

Analysis of the lifestyle questionnaire administered to 78,173 subjects (Table 3) reveals high prevalence of chronic diseases and their risk factors. Prevalence of self-reported hypertension (10.4%) and diabetes (12.1%) is higher in women compared with that in men (8.9 and 7.3%, respectively). Tobacco use (46.7%) and alcohol consumption (31.3%) are major risk factors in men but uncommon in women.

What are the main strengths and weaknesses of the study?

The PROLIFE study is in line with the WHO concept of developing non-conventional methods at the community level as a means of filling information gaps in individual countries and strengthening their health information systems. Our selection of a rural population in a rapidly urbanizing society is expected to provide new insights into the nature and consequences of epidemiological transition in health. Appropriate primordial prevention strategies can be evolved only with such knowledge. In the years
ahead, we hope to capture the dynamic relationship between conventional and newly discovered risk factors and chronic diseases. The large sample size and high prevalence of chronic diseases give us unique opportunities to explore such relationships.

One limitation of the study is that the methodology used for assigning the cause of death has not yet been validated from the conventional angle. Our validation was conducted by a single physician in 666 deaths from 1 July 2002 to 30 June 2003. For most individual causes it showed a sensitivity of ~50% and specificity of >90%. We found that sensitivity increases considerably without affecting specificity when broad categories like CVD are used. Some authors consider that 50% sensitivity and 90% specificity as reasonable in population studies.\textsuperscript{11,12} The use of semi-skilled health workers may be viewed as a weakness. However, we decided to persist with the use of health workers, as one of the principal aims of the study was to improve local level health data collection. Regular interaction with the health workers ensures that their knowledge and skill levels remain high. Refresher training programmes were conducted between December 2003 and August 2004 to improve the sensitivity of the cause of death assignment. We are planning a repeat validation study of cause of deaths and another training programme in 2009. Owing to resource limitations, all baseline risk factors have not yet been mapped. We hope to overcome this problem soon.

Can I get hold of the data? Where can I find out more?

The study’s data set is not freely available. However, we welcome specific queries and proposals for collaboration; these should be addressed to the corresponding author.

Funding

Kerala Research Program for Local Level Development (KRPLL) and HAP.\textsuperscript{1}

Acknowledgements

The authors thank the residents of Manamboor, Ottoor, Edava, Chemmaruthy, Elakamon, Cheruniyoor and Vettoor panchayats who participated in the study. They are thankful for the help provided by officials in the seven panchayats. They are also grateful to the 105 health workers and the seven supervisors.

Conflict of interest: None declared.

References

\textsuperscript{1} Health Action by People. http://www.hapindia.org (10 June 2008, date last accessed).