Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2006;35:243–249
doi:10.1093/ije/dyl042

COHORT PROFILE

Cohort Profile: The Mexico City Prospective Study

Roberto Tapia-Conyer,[1] Pablo Kuri-Morales,[2] Jesús Alegre-Díaz,[2] Gary Whitlock,[3]* Jonathan Emberson,[3] Sarah Clark,[3] Richard Peto[3] and Rory Collins[3]

How did the study come about?

The Mexico City Prospective Study (Estudio Mexicano de Cohorte Para Enfermedades Crónicas en una Población Metropolitana) is a blood-based cohort study involving follow-up of 150 000 adults (50 000 men and 100 000 women) who were aged at least 35 years when recruited in 1998–2004. The study resulted from discussions in the early 1990s about how best to measure the changing health effects of tobacco in Mexico. These discussions evolved into a plan to establish a prospective cohort study that could investigate not only the health effects of tobacco but also those of blood lipids and various other risk factors. Ideally, such a study should be big enough to assess the effects of risk factors not only overall but also in specific circumstances (e.g. at particular ages and separately in each sex) and at different levels of other risk factors (to examine possible causal interactions). In Mexico, there have been substantial decreases in overall mortality from infectious diseases over the past few decades, leaving the chronic diseases of middle age as the main causes of premature death.[1] However, the effects of risk factors for chronic diseases can depend on the presence or absence of other risk factors (including some that are measurable in blood), and these could differ greatly between Mexico and Western countries. To gain a clearer understanding of the major determinants of morbidity and premature mortality in Mexico, the Mexican Ministry of Health (Secretaria de Salud SSA) decided in 1994 to help fund a large prospective cohort study of adults in Mexico City. The Ministry staff designed the study and carried out its baseline survey, while Oxford's Clinical Trial Service Unit (CTSU), with major support from the UK Wellcome Trust for this collaboration, provided technical assistance with blood collection and storage.

What does it cover?

The chief aim of the Mexico City Prospective Study is to assess reliably the associations of established risk factors, and of possible new risks factors, with the common causes of death in Mexico. Table 1 lists Mexico's most common causes of death in 2000 by age (35–54 and 55–74) and sex. Ischaemic heart disease, stroke, diabetes (chiefly type 2) and alcoholic liver disease were major causes of death in all of these age-sex groups, as were external causes (particularly transport-related injury and homicide) in men aged 35–54, chronic obstructive pulmonary disease and lung cancer in men aged 55–74, and cervical cancer and breast cancer in women aged 35–74. Many of the deaths in which diabetes was recorded as the underlying cause are likely also to have involved ischaemic heart disease or stroke (and vice versa), so the numbers for each of these causes in Table 1 may well be underestimates. Although one of the main aims of the study is to investigate risk factors for mortality (which, indirectly, will provide information for certain causes of morbidity), the study also aims to seek information on the incidence of cancer, non-fatal vascular events, and various diseases that do not usually cause death.

Who is in the sample?

The study is set in two contiguous urban districts (Coyoacán and Iztapalapa) of central and north-east Mexico City (Figure 1). These districts contain a diverse, but settled, mix of long-term residents and relatively recent migrants from the north and south of the nation of Mexico. Chronic diseases often associated with affluence (e.g. ischaemic heart disease, and diabetes) tend to be most prevalent in the north of the country, whereas chronic diseases normally associated with poverty (e.g. stomach cancer and chronic obstructive pulmonary disease) tend to be most prevalent in the south. All men and women aged at least 35 who were living in these districts were eligible to participate in the study. The study was approved by scientific and ethics committees within the Mexican National Council of Science and Technology and the Mexican Ministry of Health.

A confidential record of all households in the two districts was compiled in 1995–97 (i.e. before recruitment in the main study started) by census-style door-to-door interviews. This record included information on the name and age of each household member, and maps showing where stairs and entrances were located. Once the main study started, the recruitment teams (each comprising two or three specially trained nurses) visited individual households, working systematically through this record. During the recruitment period, individuals in 112 333 households with eligible inhabitants were invited to take part, and one or more individuals from 106 059 of these households provided consent. The baseline assessment of each participant took place in the participant's home.

[1] Subsecretaría de Prevención y Promoción de la Salud, Direccion General de Epidemiologia, Mexico.
[2] Secretaria de Salud, Direccion General de Epidemiologia, Mexico.
[3] Clinical Trial Service Unit, University of Oxford, UK.
* Corresponding author. E-mail: gary.whitlock@ctsu.ox.ac.uk

Table 1 Leading causes of death for Mexico in 2000, by age (35–74) and sex

	Male				Female			
	Age 35–54		Age 55–74		Age 35–54		Age 55–74	
	Cause	Deaths	Cause	Deaths	Cause	Deaths	Cause	Deaths
1	Alcoholic liver disease	8176	Diabetes	11 143	Diabetes	3411	Diabetes	13 687
2	Diabetes	3701	Ischaemic heart disease	9912	Cervical cancer	1802	Ischaemic heart disease	6522
3	Ischaemic heart disease	3248	Alcoholic liver disease	7356	Breast cancer	1481	Stroke	4299
4	Transport-related injury	3200	Stroke	4460	Alcoholic liver disease	1448	Alcoholic liver disease	2674
5	Homicide	2907	COPD	2843	Ischaemic heart disease	1304	COPD	1931
6	HIV/AIDS	1572	Lung cancer	2292	Stroke	1143	Hypertensive disease	1838
7	Stroke	1276	Transport-related injury	1602	Transport-related injury	673	Cervical cancer	1746
8	Suicide	755	Prostate cancer	1445	Hypertensive disease	486	Breast cancer	1285
9	Tuberculosis	654	Hypertensive disease	1393	Chronic renal failure	450	Heart failure	1089
10	Pneumonia	623	Stomach cancer	1285	Stomach cancer	430	Chronic renal failure	1087
11	Lung cancer	495	Pneumonia	1142	Ovarian cancer	340	Lung cancer	981
12	Chronic renal failure	481	Chronic renal failure	1054	Homicide	329	Stomach cancer	946
13	Stomach cancer	480	Homicide	970	Obstetric causes	317	Pancreatic cancer	753
14	Falls	425	Heart failure	950	Leukaemia	314	Pneumonia	747
15	Drowning	424	Malnutrition	778	Lung cancer	306	Gallbladder cancer	677
16	Hypertensive disease	416	Tuberculosis	680	Tuberculosis	281	Malnutrition	624
17	Lymphoma etc.[a]	293	Pancreatic cancer	669	Rheumatic heart disease	276	Intestinal cancer	537
18	Leukaemia	289	Intestinal cancer	638	Pneumonia	272	Transport-related injury	525
19	Intestinal cancer	284	Lymphoma etc.[a]	636	Intestinal cancer	255	Ovarian cancer	520
20	COPD	282	Falls	519	Heart failure	254	Lymphoma etc.[a]	506
	Other defined causes	13 879	Other defined causes	19 809	Other defined causes	8525	Other defined causes	15 877
	Ill-defined causes	404	Ill-defined causes	738	Ill-defined causes	215	Ill-defined causes	522
	Total	44 264	Total	72 314	Total	24 312	Total	59 373

Data source: WHO mortality data.[1] National population in 2000: men aged 35–54, 9 222 304 (~1/302 were in the present study at baseline); men aged 55–74, 3 747 619 (~1/213); women aged 35–54, 10 195 983 (~1/152); women aged 55–74, 4 217 363 (~1/130).
[a] Lymphoma and other haematopoietic neoplasms excluding leukaemias. COPD = chronic obstructive pulmonary disease.

Table 2 shows characteristics of the 159 546 participants who completed the baseline questionnaire. Two-thirds were women, the mean age at survey was 53 (with 92% aged 35–74), and 60% lived in Iztapalapa. About one in seven participants had attended university or college (15.3%), and a similar proportion (13.5%) had never received any formal education. A total of 37 433 people reported a history of one or more specific chronic diseases, including 21 810 people (13.7% of all participants) who reported a history of diabetes, 2503 (1.6%) a history of ischaemic heart disease ('heart attack' or angina), and 2023 (1.3%) a history of cancer (any site).

How often are they followed up, and what is the rate of loss likely to be?

Participants are being followed up indefinitely for cause-specific mortality through Mexican death registries. In Mexico, almost all adult deaths are certified by a doctor, and the underlying and contributing causes of death are coded according to the 10th International Classification of Diseases. This information, together with personal identifying information (e.g. name, dates of birth and death, addresses of residence and death), is recorded in a Ministry of Health database that was established as a consequence of this study (as well as in other official databases). The investigators based in the Ministry of Health have regular confidential access to this database and use the personal identifying information to link death records automatically to study participants (typically within 2 or 3 months of a participant's death). Paper copies of death certificates can be obtained for central review (as will be done, for example, for all deaths involving diabetes, ischaemic heart disease, or stroke as an underlying cause: see above), and individual households can be revisited to verify vital status. Any participant who emigrates abroad is likely to be lost to mortality follow-up, but it is estimated that this would apply to much <1% of study participants annually. Slight losses to follow-up might also occur if people move out of Mexico City into regions where death certification is not yet complete. But, even in rural regions, an average of 98% of adult deaths are certified by a doctor and so will be linked automatically.

At 5 year intervals, a reasonably representative sample of at least a few thousand surviving participants will be invited for re-assessment, including the same questions, measurements, and blood collection procedures (described below) that were used at baseline. These repeat assessments will be used to help take account of biological variation and random errors in measurements made at baseline.[2]

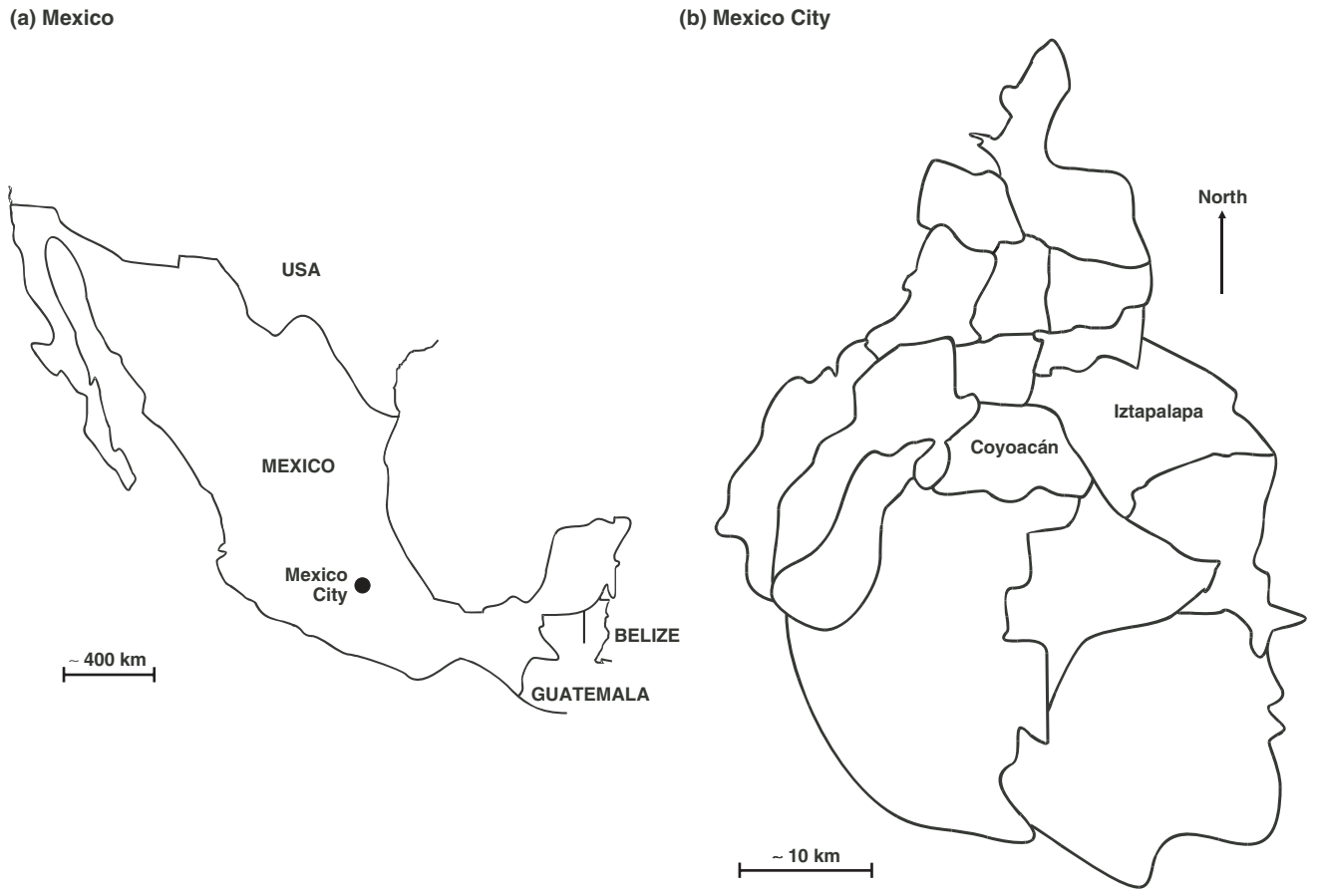

Figure 1 Maps showing the locations of (a) Mexico City within Mexico, and (b) the Coyoacán and Iztapalapa districts within Mexico City

Table 2 Baseline characteristics (unadjusted) of study participants

	Men, by age (years)			Women, by age (years)			
	35–54	55–74	⩾75	35–54	55–74	⩾75	**Total**
Number of participants	30 553	17 559	4472	67 004	32 328	7630	159 546
Age, years, mean	43.8	63.4	80.4	43.6	63.2	80.8	52.6
Resident in Coyoacán (%)	44.7	42.2	35.1	38.7	38.6	35.6	40.0
Resident in Iztapalapa (%)	55.3	57.8	64.9	61.3	61.4	64.4	60.0
Highest educational attendance							
University/college (%)	31.7	12.8	6.8	15.8	4.4	2.3	15.3
High school (%)	33.0	16.3	8.9	31.5	11.3	6.0	24.2
Elementary school (%)	32.1	55.8	50.9	46.1	57.8	45.1	47.0
Other (%)	3.2	15.1	33.3	6.6	26.4	46.4	13.5
History of one or more serious chronic diseases[a] (%)	12.9	31.8	34.0	16.6	38.5	37.0	23.5
Diabetes (%)	8.6	21.6	17.9	7.6	24.5	21.1	13.7
Ischaemic heart disease (%)	0.8	3.6	5.5	0.6	2.2	3.9	1.6
Stroke (%)	0.5	2.1	3.6	0.6	1.8	3.0	1.2
Cancer (any site) (%)	0.3	1.0	2.5	1.2	2.1	2.3	1.3
Other (%)	3.5	7.7	10.6	7.8	13.6	12.4	8.5
Blood sample in storage (%)	97.2	97.0	97.2	97.8	97.4	96.5	97.5

[a] Self-reported personal history of diabetes mellitus (type unspecified), ischaemic heart disease ('heart attack' or angina), stroke, peripheral arterial disease, cancer (any site), emphysema, chronic kidney disease, peptic ulcer or cirrhosis.

What has been measured?

For a large study to be practically and economically feasible, its procedures need to be simple and streamlined. In this study, the full recruitment assessment (including obtaining informed consent and collecting blood) typically took just 30–40 min to complete, and it was nearly paperless (in common with the recently described Kadoorie Study[3]). As the information was being obtained, data from a simple interviewer-administered questionnaire (summarized in Table 3) and physical measurements (summarized in Table 4) were entered directly into a handheld data recorder (of one of two types, described in Table 5). At the end of each working day, this information was automatically uploaded to the study's electronic database. These electronic procedures minimized the costs of data collection, processing, and checking, while improving data accuracy and completeness.

Table 3 Questionnaire data collected in the Mexico City Prospective Study

Demographic data
Socioeconomic status
• Education
• Occupation
• Income
• Home ownership
Personal health behaviours
• Alcohol
• Smoking
• Fruit and vegetable consumption
• Fried food consumption
• Cooking oils
• Recreational physical activity
• Sleep habits
General health-related data
• Disease history (for 18 common conditions)
• Medications
Reproductive history (for women)
• Contraceptives
• Pregnancies
• Hysterectomy
• Ovarectomy

Information was entered into the handheld recorders either by a touch-sensitive screen or by a combination of barcode and alphanumeric keypad. The handheld recorders had been programmed to take the field workers through the questionnaire in the same prescribed manner, automatically making appropriate skips (e.g. past questions about obstetric history if the participant was male), and enabling the field workers to modify previous entries or to void the questionnaire entirely if necessary. In addition, the data recorders were programmed to query moderately extreme physical measurements and to prevent highly implausible measurements from being recorded. After the data had been uploaded to the study database, and multiple backup copies made, a database-checking programme was used to identify possible data errors. Such errors were reviewed regularly by a data monitoring team and, when necessary, checked by field workers (e.g. by revisiting participants). However, because of the in-field automated checks of unusual values, there was less need for subsequent checking than originally expected.

The blood sample (10 ml, venous) was collected into a single EDTA tube labelled with a barcode unique to the participant. This barcode was scanned, and pertinent information (e.g. date and time of collection, and time of last meal) was entered into the handheld recorder. The sample was then placed in an insulated box containing several chilled packs, reliably maintaining an internal box temperature of 4–10°C. At the end of the day, the boxes were transported to a central laboratory in Mexico City where the samples were extracted and placed in a refrigerator (4°C) for overnight storage. The next morning the samples were centrifuged (2100*g* at 4°C for 15 min) and then the plasma was taken off and divided equally between two 1.8 ml cryovials (Nunc A/S, Denmark), and the buffy coat (the layer, rich in white blood cells, between plasma and red blood cells) was transferred to a third cryovial of the same type. Each cryovial in a set of three was pre-printed with a barcode unique to the particular set. This barcode, together with the barcode on the original blood collection tube, was scanned into a computerized inventory program to establish a unique link between them. The three cryovials were placed in identical positions in three separate storage boxes. These boxes were stored in three separate −80°C freezers (to prevent complete loss of an individual's samples in the event of a breakdown) until freighted by air in large insulated boxes (each containing 30 storage boxes and 40 kg of dry ice, which had been shown in pilot studies to keep specimens frozen for at least 5 days) to Oxford about once every 6 weeks. Upon receipt

Table 4 Clinical measurements at baseline survey in the Mexico City Prospective Study

Variable	Number of measurements	Equipment used	Precision
Standing height	1	Wooden triangle + 3 m long measuring tape	Nearest millimetre
Waist circumference	1	3 m long measuring tape	Nearest millimetre
Hip circumference	1	3 m long measuring tape	Nearest millimetre
Weight	1	Portable analogue scales[a] (1998–2003)	Nearest kilogram
		Portable electronic scales[a] (2003–04)	Nearest 100 g
Resting pulse rate	3 (each 3 min apart)	Manual palpation	Beats per minute
Sitting blood pressure	3 (each 3 min apart)	Standard mercury sphygmomanometer[a]	Nearest 1 mm Hg

[a] Scales recalibrated twice a week, sphygmomanometer recalibrated once a month.

Table 5 Handheld data recorders used in the baseline interviews

Model	Period used	Features
Microwand 32ES barcode optical reading device (Hand Held Products)	April 1998–November 2001	Most questionnaire information entered by barcode; alphanumeric keyboard for other information (e.g. anthropometric measurements)
SPT 500 personal digital assistant with palm operating system (Symbol)	November 2001–December 2004	All questionnaire information entered by touch-sensitive screen; could store 70 completed questionnaires

at the CTSU laboratory in Oxford (usually within 2–3 days of dispatch) the small storage boxes were immediately unpacked and placed in tanks containing liquid nitrogen vapour at −150°C. A PC-based inventory system developed for the study records the storage location of cryovials within boxes and of boxes within tanks.

Blood is now in storage for 155 487 participants (i.e. 97.5% of those with a completed questionnaire: Table 2). The mean delay (at 4–10°C) before blood separation was ~24 h. Previous research has shown lack of significant effects for many analytes with delays of separation of more than 2–3 days, even at room temperature.[4,5]

What has the study found?

Analyses of associations between possible risk factors and cause-specific mortality will begin in a few years when sufficient numbers of cause-specific deaths have occurred. Disentangling deaths that involve diabetes, ischaemic heart disease, or stroke will require inspection of death certificates (see above); however, among those aged <75 at baseline, some 2000 deaths involving at least one of these causes could (from WHO mortality data: e.g. see Table 1) be expected within the first 5 years of follow-up (i.e. by 2007–08) and, perhaps, some 4000–6000 such deaths within the first 10 years (as well as 300–400 deaths each from cancer of the lung, intestine, cervix, and breast).

In addition to information on possible risk factors collected at baseline (Tables 3 and 4), possible risk markers in blood—potentially including lipid subfractions, apolipoproteins, coagulation and inflammatory factors, hormones, vitamins, markers of insulin resistance, and genetic polymorphisms—will also be investigated. Analyses of possible blood markers for risk will use a nested case–control approach, whereby the stored blood from people who have died from a particular disease will be retrieved and compared with blood retrieved from otherwise similar individuals who do not have the disease. This approach is economical (because only a sample of the total number of specimens is assayed, and because the real cost of assays tends to diminish with time), while avoiding the potential problem of reverse causality (because in nearly all instances the blood will have been collected before the onset of disease). Moreover, it allows factors that only in the future may come to be regarded as potentially important also to be assayed.

Because the study sample is largely representative of the adult population in Mexico City, the baseline prevalences of disease (Table 2) and of disease risk factors (Table 6) are of interest for assessing local health needs and for disease prevention. The prevalence of current smoking was strongly related to age, with about one in two men (50.7%) and one in four women (25.0%) at ages 35–54 being current smokers, compared with just one in five men (19.8%) and one in 20 women (5.2%) aged ⩾75 being current smokers. Just over one-third of the men aged 35–54 (35.6%) drank alcohol at least once a month, but the proportion of men aged ⩾75 who did so was much lower (15.8%); and, in each age group, only one-quarter to one-fifth as many women as men drank this frequently. In each age group, about one in four men and one in five women reported having recreational physical activity at least once a week.

The mean BMI overall was 27.9 kg/m^2 in men and 29.5 kg/m^2 in women, but in both sexes, BMI tended to be lower at ages ⩾75 than at younger ages (Table 3). Among those aged 50–59, the prevalence of a BMI of ⩾30 kg/m^2 was 30% in men (which is slightly less than the prevalence of 31% among US men at the same age and in about the same years[6]), but it was 47% in women (or much higher than the corresponding prevalence of 37% among US women[6]).

What are the main strengths and weaknesses?

This study is probably the largest blood-based prospective cohort study to have been established so far in Latin America. It should, therefore, be well placed to address uncertainties about disease causation in that part of the world, especially since it has blood stored for nearly all of its participants. But factors that will limit the accrual of deaths, and, hence, the ability to investigate causes of death, include a fairly young age distribution (median 50 years) and an unintentionally low percentage of men (33%). Recruiting men turned out to be harder than expected because a higher proportion of men were working long hours during the day in the main study than in two earlier pilot studies, a problem that made it necessary to arrange many repeat visits during the evening and weekends. Other major challenges to recruitment included a series of hurricanes and tropical storms that hit Mexico in 2002–03, creating national and regional public health disasters to which study staff (as employees of the Ministry of Health) were seconded (partly accounting for the flat part at the right-hand end of the study's recruitment graph: Figure 2). Nonetheless, the study should still be able to address many important questions about disease causation in Mexico, and much of the information could also be relevant to other countries. For example, the high prevalence of obesity and diabetes may mean that this study can provide insights into the future impact of these conditions in countries where the epidemics of obesity and diabetes are currently less mature but which could have similarly high prevalences of these conditions within a decade or two (e.g. many Western countries).

This study was affordable only because it used simple and streamlined data and blood collection methods. For example, the handheld data recorders allowed rapid collection of complete information, avoided the need to process and store

Table 6 Baseline distributions (unadjusted) of known and possible risks factors for disease

	Men, by age (years)			Women, by age (years)			
	35–54	55–74	⩾75	35–54	55–74	⩾75	**Total**
Number of participants	30 553	17 559	4472	67 004	32 328	7630	159 546
Tobacco smoking							
Current smoker (%)	50.7	34.9	19.8	25.0	11.5	5.2	27.2
Ex-smoker (%)	28.8	45.3	56.0	17.3	17.9	18.6	23.9
Never smoked (%)	20.5	19.8	24.3	57.7	70.5	76.2	48.9
Alcohol drinking							
Daily or most days (%)	2.0	3.6	4.5	0.2	0.5	1.1	1.2
1–4 days a week (%)	13.8	10.2	5.9	1.9	1.7	1.3	5.1
1–3 days a month (%)	19.8	12.9	5.4	5.3	2.9	1.6	8.2
1–11 days a year (%)	45.5	43.7	35.3	58.9	50.0	35.7	51.1
Ex-drinker (%)	13.1	23.2	37.5	10.0	15.0	19.8	14.3
Never drank (%)	5.7	6.4	11.4	23.7	29.9	40.4	20.1
Fried food, days/week, mean (SD)	2.7 (1.9)	2.4 (1.9)	2.1 (1.8)	2.3 (1.7)	2.0 (1.7)	1.9 (1.7)	2.3 (1.8)
Fruit and veg., days/week, mean (SD)	4.1 (1.9)	4.3 (1.9)	4.4 (1.9)	4.6 (1.7)	4.7 (1.8)	4.7 (1.8)	4.5 (1.8)
Recreational physical activity							
⩾3 times a week (%)	14.7	19.7	18.8	12.9	14.8	11.0	14.5
1–2 times a week (%)	13.0	6.2	3.6	4.0	5.1	3.9	6.2
Some but < once a week (%)	3.9	1.4	0.8	0.8	0.9	0.7	1.5
None (%)	68.3	72.7	76.9	82.3	79.3	84.4	77.9
BMI, kg/m^2, mean (SD)	28.1 (4.4)	27.9 (4.2)	26.4 (3.9)	29.5 (5.3)	29.9 (5.3)	27.5 (5.0)	29.0 (5.1)
BMI ⩾ 30 kg/m^2 (%)	28.8	27.1	16.9	40.9	45.3	28.5	36.8
Waist circ, cm, mean (SD)	95.6 (10.6)	97.7 (10.5)	96.9 (10.6)	91.8 (12.0)	96.4 (12.0)	95.7 (11.8)	94.4 (11.8)
Waist:hip ratio, mean (SD)	0.94 (0.06)	0.97 (0.07)	0.97 (0.07)	0.87 (0.07)	0.90 (0.07)	0.93 (0.08)	0.90 (0.08)
SBP, mm Hg, mean (SD)	125.2 (13.6)	133.7 (17.3)	135.9 (18.3)	122.4 (14.7)	134.4 (17.9)	138.8 (19.7)	127.8 (16.9)
DBP, mm Hg, mean (SD)	83.7 (9.6)	85.9 (10.4)	84.0 (10.6)	81.1 (9.8)	85.3 (10.4)	84.4 (11.3)	83.2 (10.3)

Fruit & veg. = fruit and vegetables; waist circ. = waist circumference; BMI = body mass index; SBP = systolic blood pressure; DBP = diastolic blood pressure.

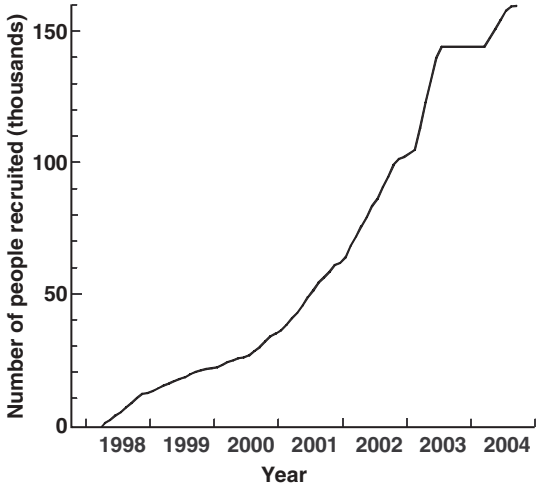

Figure 2 Recruitment of participants to the Mexico City Prospective Study

more than one million individual sheets of paper, and are expected to reduce the medium-term data collection and storage costs by about three-quarters. Paperless data collection methods such as these (including laptop- or internet-based methods[3]) have great potential for reducing costs while preserving or improving data accuracy.

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

The study's data are not freely available, but specific proposals for future collaboration, addressed (in English or Spanish) to any of the study's Mexico-based or Oxford-based investigators, would be welcomed.

Acknowledgements

We thank the following people for their help in establishing the study: David Nava Menchaca, Martin Radley, Raul Ramirez Reyes, Fernando Velazquez-Gonzalez, Clementina Magos-López, and Linda Youngman. The study has been funded by the Secretaria de Salud SSA (Mexican Ministry of Health), the Consejo Nacional de Ciencia y Tecnología (Mexican National Council of Science and Technology), and the UK Wellcome Trust. The Clinical Trial Service Unit receives core funding from the Medical Research Council, the British Heart Foundation and Cancer Research UK, and additional support for international studies of smoking from the US NIH Fogarty International Center.

References

[1] World Health Organization Statistical Information System. Available at: http://www3.who.int/whosis/menu.cfm?path=whosis,mort (Accessed January 2006).

[2] Clarke R, Shipley M, Lewington S *et al*. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *Am J Epidemiol* 1999;**150:**341–53.

[3] Chen Z, Lee L, Chen J *et al*. Cohort profile: The Kadoorie Study of Chronic Disease in China. *Int J Epidemiol* 2005;**34:**1243–49.

[4] Clark S, Youngman LD, Palmer A *et al*. Stability of plasma analytes after delayed separation of whole blood: implications for epidemiological studies. *Int J Epidemiol* 2003;**32:**125–30.

[5] Clark S, Youngman LD, Chukwurah B *et al*. Effect of temperature and light on the stability of fat-soluble vitamins in whole blood over several days: implications for epidemiological studies. *Int J Epidemiol* 2004;**33:**518–25.

[6] World Health Organization. Available at: http://www.who.int/ncd_surveillance/infobase/web/surf2/country_list.html (Accessed January 2006).