European Journal of Cancer Prevention 2001, **10**, 489–499

The Malmö diet and cancer study: representativity, cancer incidence and mortality in participants and non-participants

J Manjer[1], S Carlsson[1], S Elmståhl[1], B Gullberg[1], L Janzon[1], M Lindström[1], I Mattisson[1,2], G Berglund[1]

(Received 8 March 2001; accepted 16 July 2001)

In order to investigate potential selection bias in population-based cohort studies, participants ($n = 28\,098$) and non-participants ($n = 40\,807$) in the Malmö Diet and Cancer Study (MDCS) were compared with regard to cancer incidence and mortality. MDCS participants were also compared with participants in a mailed health survey with regard to subjective health, socio-demographic characteristics and lifestyle. Cancer incidence prior to recruitment was lower in non-participants, Cox proportional hazards analysis yielded a relative risk (RR) with a 95% confidence interval of 0.95 (0.90–1.00), compared with participants. During recruitment, cancer incidence was higher in non-participants, RR: 1.08 (1.01–1.17). Mortality was higher in non-participants both during, 3.55 (3.13–4.03), and following the recruitment period, 2.21 (2.03–2.41). The proportion reporting good health was higher in the MDCS than in the mailed health survey (where 74.6% participated), but the socio-demographic structure was similar. We conclude that mortality is higher in non-participants than in participants during recruitment and follow-up. It is also suggested that non-participants may have a lower cancer incidence prior to recruitment but a higher incidence during the recruitment period.
© 2001 Lippincott Williams & Wilkins

Key words: Cancer incidence, mortality, non-participants, participants, representativity, socio-demography.

Introduction

In order to evaluate the results from population-based prospective cohort studies (e.g. disease and mortality rates and the prevalence of different risk factors) it is important to identify differences between participants and non-participants. This includes two perspectives. The first is related to differences at baseline (e.g. prevalent disease, socio-demographic characteristics and lifestyle factors). The second issue concerns differences with regard to disease rates and mortality during follow-up.

The Malmö Diet and Cancer Study (MDCS) is a population-based prospective cohort study that between 1991 and 1996 recruited 44- to 74-year-old men and women living in Malmö, a city in southern Sweden with about 250 000 inhabitants. The main goal of the MDCS is to study the impact of diet on cancer incidence and mortality (Berglund, 1993). It consists of a baseline examination including dietary assessment, a self-administered questionnaire, anthropometric measuring and collection of blood samples that have been stored in a biological bank (Berglund *et al*., 1993). Follow-up is performed regularly by means of record linkage with national registries for mortality and cancer incidence.

The aim of the present analysis was to evaluate the representativity of the study group by analysing cancer incidence and mortality prior to invitation (for cancer incidence), during recruitment and following end of baseline examination, in participants and non-participants. An additional aim was to compare subjective health, socio-demographic factors and lifestyle in MDCS participants with a random sample of subjects from corresponding birth cohorts in Malmö who participated in a mailed health survey with high participation. This study will also discuss the potential impact an unrepresentative sample may

[1]Department of Community Medicine; [2]Department of Medicine, Lund University, Malmö University Hospital, SE-205 02 Malmö, Sweden. Correspondence to: G Berglund. Fax: (+46) 40 92 32 72.

have on different types of studies (i.e. studies on disease and exposure rates versus studies on association between exposure and disease).

Materials and methods

The Malmö Diet and Cancer Study

The Malmö Diet and Cancer Study (MDCS) was shaped during discussions in the early 1980s between representatives for the International Agency for Research on Cancer (IARC), The Swedish Cancer Society, the Swedish Medical Research Council and the Faculty of Medicine, Lund University, Sweden. The main objective of the MDCS is to clarify whether a Western diet high in fat and total calories while low in vegetables, fruit and fibres, increases the risk of certain forms of cancer such as cancer of the breast, colon, rectum, pancreas, ovary, endometrium and prostate (Berglund, 1993).

A novel dietary assessment method was developed and the validity (Riboli *et al*., 1997) and reproducibility (Elmståhl *et al*., 1996) of that method was tested. A pilot study with the aim of testing the chain of events for recruiting was performed in 1990 and baseline examination started in March 1991. The design included the creation of a versatile biological bank with separated blood components from all participants (Berglund *et al*., 1993). The quality of the frozen blood components over time has been presented in a quality control programme (Pero *et al*., 1998a,b). In 1993 the MDCS became an associated member of the European Prospective Investigation into Cancer and Nutrition (EPIC) organized by the International Agency For Research on Cancer (IARC), WHO, Lyon, France (Riboli, 1992).

Recruitment

All Swedish residents are registered in the National Population Register where they are given a unique civil registration number. All subjects living in Malmö 1 January 1991 and born 1926–1945 were included in the first group that was invited to the MDCS, a total of 53 325 individuals. Following this, the cohort was redefined every third month, including people that had moved to Malmö and identifying diseased subjects and individuals who had moved from the city, which guaranteed that ineligible subjects could be removed from the sampling frame rather than classified as non-responders. The study was extended in 1995 to include men born between 1923 and 1945 and women born 1923–1950. The rationale of including more young women than men, was to be able to study breast cancer among premenopausal women. Recruitment was carried out between 1 January 1991 and 25 September 1996. There were in all 74 138 individuals in the selected birth cohorts according to the population register (Figure 1). Seventeen subjects could not be satisfactorily identified by their civil registration number, some of them were individuals who had been assigned a civil registration number twice (e.g. had immigrated to Sweden more than once) (Figure 1).

The background population was randomly invited by letter. If a subject did not answer the first letter, two more letters were sent. In parallel to invitation by personal letter, subjects living in Malmö from the studied birth-year cohorts were invited by advertisement in local newspapers, in public places and in primary health care centres. In this way, a total of 5505 individuals were recruited without having received a letter of invitation. These were referred to as 'volunteers' (Figure 1). Most volunteers were recruited during the first half of the study.

Some individuals died or moved before they had received a letter of invitation: in all 3017 subjects (Figure 1). There were 21 817 subjects who did not reply to the invitation letters (Figure 1). A smaller number, 16 942 individuals, answered the letter but were not willing to take part in the study. A minor group, 224 subjects, came to the first meeting but died before they had completed baseline examination (Figure 1). The only exclusion criteria that were decided on a priori were language problems and mental retardation that prevented the respondent from answering the rather extensive questionnaire (Berglund *et al*., 1993). This group eventually included 1975 individuals (see below). A group of 2048 persons began baseline examination but failed to complete either the questionnaire, the anthropometric/body composition measurements or the dietary assessment (Figure 1).

A subject who completed the questionnaire, the anthropometric/body composition measurements and the dietary assessment was regarded as a complete participant. This information is a requirement for eligibility to the EPIC study (Riboli and Kaaks, 1997). The cohort of 28 098 complete participants, 11 063 men and 17 035 women, will be referred to as the EPIC cohort. Some subjects were not considered eligible for study (i.e. those who could not be satisfactorily identified by their civil registration number, those who died or moved prior to invitation or died following their first visit and before complement, and individuals that had language problems or were mentally retarded). The rest of the background

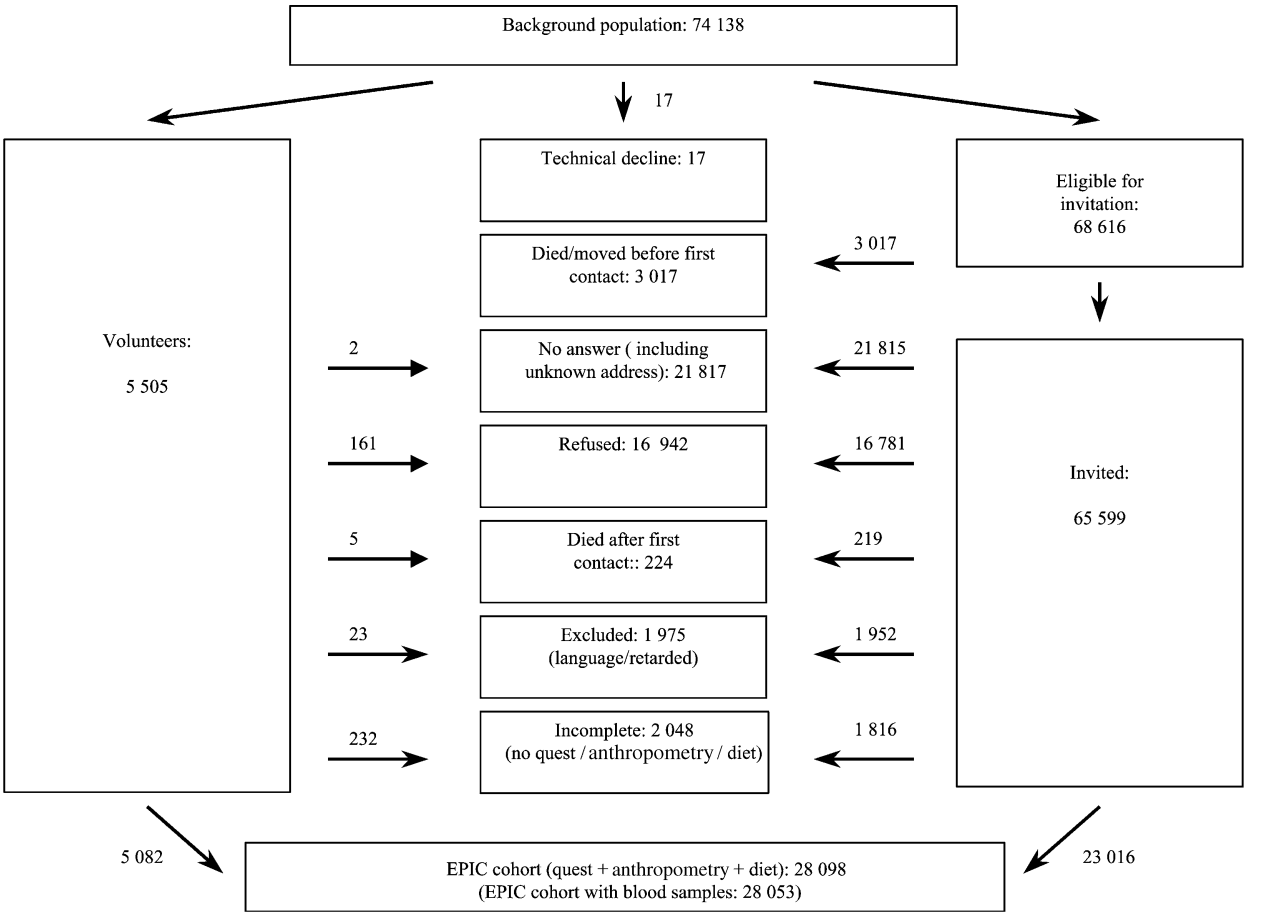

Figure 1. Recruitment and participation in the Malmö Diet and Cancer Study.

population will be referred to as non-participants or the non-EPIC cohort, in all 40 807 subjects, 17 810 men and 22 997 women. This resulted in an overall participation rate of 40.8%, for men 38.3% and for women 42.6%. Mean age at 1 January 1991 was 54.9 (SD 7.6) years for participants and 54.3 (SD 7.8) years for non-participants.

Baseline examination

Subjects were invited to a first visit where the aim of the study and the different parts of the baseline examination were presented. A self-administrated questionnaire assessed education, occupation, physical activity, social network, use of tobacco and alcohol, current health, medical history, current medication and disease in close relatives. Women were also asked about reproductive history. The questionnaire was handed out at the first visit and was collected and checked for missing values at a second visit, typically 2 weeks later. Dietary assessment was performed according to a modified diet history method, which had been validated (Riboli *et al*., 1997). Participants completed a menu book for 7 consecutive days and a food frequency questionnaire, which were discussed and completed at the second visit. At the first visit, blood pressure, height, weight and body composition were measured, the latter by an impedance method (Berglund *et al*., 1993). Samples of 45 ml blood were drawn which were separated and stored in the biological bank at −80°C or −140°C according to blood component fraction. Some participants were randomly selected for additional studies such as ultrasonographic examination of carotid arteries (Persson *et al*., 1997) and bone density measurements.

Subjective health, socio-demographic factors and lifestyle

An attempt was made to assess differences between participants and non-participants with regard to es-

tablished and potential risk factors for cancer incidence and mortality. As no such information is available considering non-participants, participants in the MDCS were compared with subjects in Malmö who took part in a mailed health survey with an almost twice as high participation rate in corresponding age groups, The Health Situation in Malmö '94 (HSM:94) (Lindström *et al*., 1995). The MDCS questionnaire was mailed in 1994 to a random sample of subjects from some birth-year cohorts also included in the MDCS (i.e. 1923, 1933 and 1943). In all, 2100 subjects were selected in these birth cohorts in the HSM:94 and 1567 returned the questionnaire, resulting in a response rate of 74.6%. Selection of eligible subjects for the HSM:94 was made without any relation to participation in, or invitation to, the MDCS. In the same birth cohorts in the MDCS there were 8331 eligible subjects out of whom 3196 completed baseline examination, corresponding to a participation rate of 38.4%. The distributions of some selected health, socio-demographic and lifestyle factors were compared between the EPIC and the HSM:94 cohorts. The question on education included an open alternative in the HSM:94 questionnaire, and the distribution when this alternative was excluded has also been reported. Type of occupation was used to define those currently employed, the only group included in the comparison of type of employment and current sick leave. Moreover, the latter factors were only compared in subjects born in 1933 and 1943, as most participants born in 1923 were retired at baseline examination. Marital status did not cover whether a person was cohabiting or not but this was assessed in the 'living alone' question. Smoking habits and alcohol consumption were categorized according to Table 1. Weight status (body mass index, kg/m^2) was assessed by direct measurement at baseline examination in the MDCS cohort while it was based on self-reported information in the HSM:94. Subjective health was measured on a scale with seven fixed steps spanning from 'very bad, could not be worse' to 'very good, could not be better'. The assessment of instrumental support and social participation have been described in detail elsewhere (Hanson *et al*., 1997).

Follow-up on cancer

We have used the EPIC definition of a tumour end point that states 'all malignant tumours and certain tumours with non-malignant behaviour' (EPIC, 1998). The latter include: 'carcinoma *in situ* of the breast, carcinoma *in situ* of cervix uteri, all neoplastic changes in the urinary bladder and all neoplasia in the brain'.

Tumour end points were retrieved by record linkage with the Swedish Cancer Registry and the Southern Swedish Regional Tumour Registry. The Swedish Cancer Registry was set up in 1958 (National Board of Health and Welfare, 1998a). By law, all malign tumours and certain benign tumours are to be reported, this includes all malignancies mentioned above. The national register is complete with a two-year delay, in this study until 31 December 1997. Additional information until end of follow-up, 31 December 1999, was obtained from the Southern Swedish Regional Tumour Registry which provides up-to-date information on cancer incidence in the south of Sweden (Southern Swedish Regional Tumour Registry, 1994).

Tumour site has been registered according to the ICD version that was used at diagnosis. In addition, to facilitate comparisons over time, all tumours are also coded according to ICD-7. Since 1958, histopathological type has been registered using the C24 classification (National Board of Health and Welfare, 1998a). SNOMED and ICDO-2 code is available on all cases diagnosed after 1 January 1993, but in order to enable international comparisons, all tumours in the registry have been coded retroactively according to these classifications.

Follow-up on vital status

The present cause-of-death register was set up in 1911. All deaths, immediate cause-of-death and multiple causes of death are to be reported (National Board of Health and Welfare, 1998b). Information on cause-of-death is normally available with a one-year lag, but due to a delay caused by the introduction of the ICD-10 classification, cause of death and information on the use of autopsy was only available up until 31 December 1997. The cause-of-death registry was completed until 31 December 1999 (i.e. end of follow-up), with data from the National Tax Board that provides up-to-date information on vital status for all Swedish residents.

Statistical methods

The distributions of selected factors in the EPIC and HSM:94 cohorts were compared using direct age standardization. Age-standardized rates were calculated as the sum of birth-years cohort-specific rates multiplied by a third (or by a half for type of employment and current sick leave as these factors

Table 1. Comparison between an age-matched subset of the EPIC-cohort and participants in the MHS:94

Factor group	EPIC cohort % with factor,[a]	HSM:94	
Born in Sweden	89.1	74.9	
Education			
O-level college (≤ 9 years)	69.5	61.8	68.7
A-level college (≤ 12 years)	8.7	7.2	7.9
University/university college	21.6	19.5	21.1
Other (only in HSM:94)	–	9.4	(excluded)
Type of occupation			
Domestic work	1.5	2.7	
Employed	51.5	41.2	
Retired	42.1	44.5	
Student	0.3	0.6	
Unemployed	4.3	5.4	
Type of employment (in currently employed)			
Manual	32.2	35.5	
Non-manual	56.0	54.0	
Self-employed − employer	11.8	6.9	
Marital status			
Married	64.8	64.5	
Unmarried	9.1	9.1	
Divorced	15.7	15.9	
Widowed	10.3	9.2	
Living alone	27.3	28.5	
Smoking			
Never	39.9	37.5	
Current	25.5	27.2	
Ex	34.6	31.0	
Alcohol			
Nothing last year	10.1	11.7	
Something last year but not last month	11.2	11.1	
Something last month	78.7	57.6	
Weight (BMI)			
Normal (< 25)	44.7	47.1	
Overweight (25−30)	41.0	38.6	
Obese (≥ 30)	14.0	11.0	
Poor well-being (1−2 out of 7)	2.8	8.3	
Current sick leave (in currently employed)	5.5	4.1	
High instrumental support (would certainly get practical help from someone if needed)	69.8	61.0	
High social participation (went to 3 or more out of 13 selected activities last year, e.g theatre, party)	66.5	53.1	

[a] With missing answers included in denominator (i.e. do not always add to 100%).

were only examined in subjects born in 1933 and 1943).

Incidence of first time malignancy, the studied end point being any malignancy according to the EPIC definition, was compared between EPIC and non-EPIC cohorts. Cancer incidence was studied in relation to the recruitment period, which resulted in three different periods. The first from the establishment of the Swedish Cancer Registry, 1 January 1958, until definition of the original cohort, 31 December 1990. The second from definition of the original cohort, 1 January 1991, until end of baseline examination, 25 September 1996. The third period following end of baseline examination, 26 September 1996, until end of follow-up, 31 December 1999. Mortality in EPIC and non-EPIC cohorts was studied for the last two periods. Incidence of specific cancer sites and cause specific mortality was studied in participants and non-participants for the periods 1 January 1991 to 31 December 1999 and 1 January 1991 to 31 December 1997, respectively.

Incidence and mortality rates were calculated as the number of events per 100 000 person-years. Person-years were counted from beginning of each period until first event, death or end of the period. First time malignancies were only counted once (e.g. a subject with a colorectal carcinoma in 1965 was censored from further analysis with regard to first time malignancy). Correspondingly, in the analysis of specific carcinomas, each site was only counted once

and the subject was censored from further analysis of this site (e.g. a person with a colorectal carcinoma in 1965 was censored from further analysis of colorectal carcinoma but not from the analysis of other sites).

Incidence and mortality rates were compared between the EPIC and non-EPIC cohorts using Cox's proportional hazards analysis adjusted for sex and age at definition of the original cohort, 1 January 1991. Analyses of first time malignancy and all cause mortality were repeated stratified on sex.

Results

Subjective health, socio-demographic factors and lifestyle (Table 1)

The high response rate, 74.6%, in the HSM:94 implies that it might be more representative of the background population than is the EPIC cohort. The distributions were similar in the two cohorts with regard to educational level, type of employment, marital status, percentage living alone, smoking habits, weight distribution and current sick leave. There were fewer foreigners in the EPIC cohort than in corresponding birth-year cohorts in the HSM:94. The EPIC cohort had a relatively higher proportion of employed and a lower proportion of retired subjects than had the HSM:94 cohort in corresponding birth-year cohorts. The question on alcohol had many missing values in the mailed survey, 20%, which makes comparison difficult, but the proportion of teetotallers was similar in both cohorts. A relatively higher proportion of subjects in the EPIC cohort reported good subjective health, good instrumental support and high social participation. Taken together, the data suggest that the EPIC cohort is selected towards better health but that the socio-demographic structure and the prevalence of smoking and obesity are the same as in a study, from the same population, with a participation rate of 74.6%.

Cancer incidence (Tables 2 and 3)

Incidence of first time malignancy in the period preceding definition of the original cohort was lower in non-participants than in participants, relative risk (RR): 0.95 (0.90–1.00), but stratified for sex, this was conferred to women only, RR: 0.94 (0.89–0.99). Cancer incidence was higher in the non-EPIC cohort during recruitment, 1.08 (1.01–1.17), but the association did only reach statistical significance for men, 1.14 (0.1.02–1.28). This implies that subjects with previous cancer diagnosis may have been more willing to participate and that non-participants may be at higher risk of developing cancer during the recruitment period. Following recruitment, both men and women in the non-EPIC cohort had a similar incidence of first time malignancies as subjects in the EPIC cohort.

From definition of the original cohort until end of follow-up, some cancer sites were significantly more common in non-EPIC subjects, namely: tumours of the pancreas, lung, cervix uteri (invasive) and the urinary tract (Table 3). The EPIC-cohort had a higher incidence of breast cancer, prostate cancer, and malignant melanoma of the skin.

Mortality (Tables 4 and 5)

During recruitment and follow-up, 982 subjects in the EPIC cohort died, 293 during and 689 following end of recruitment. Out of these, 527 were diagnosed prior to 31 December 1997 (i.e. during the period when information on cause of death and

Table 2. Incidence of first-time malignancy in EPIC and non-EPIC cohorts

Period	Group	No. of individuals	No. of first time malignancies	Incidence/10^5	RR[a] (95% CI)	Sex-specific RR[b] (95% CI)	
						Men	Women
1 January 1958 to	EPIC cohort	28 098	2470	276	1.00	1.00	1.00
31 December 1990	Non-EPIC	40 807	3213	246	0.95 (0.90–1.00)	1.02 (0.89–1.19)	0.94 (0.89–0.99)
	Total	68 905	5683	258			
1 January 1991 to	EPIC cohort	25 628	1099	764	1.00	1.00	1.00
25 September 1996	Non-EPIC	37 593	1684	805	1.08 (1.01–1.17)	1.14 (1.02–1.28)	1.05 (0.95–1.16)
	Total	63 221	2783	788			
26 September 1996 to	EPIC cohort	24 357	1013	1310	1.00	1.00	1.00
31 December 1999	Non-EPIC	34 694	1387	1278	1.00 (0.93–1.09)	1.06 (0.94–1.19)	0.96 (0.86–1.08)
	Total	59 051	2400	1292			

[a]Adjusted for sex and age at definition of original cohort, 1 January 1991.
[b]Adjusted for age.

Table 3. Incidence of specific carcinomas in EPIC and non-EPIC cohorts. From definition of original cohort until end of follow-up (1 January 1991 to 31 December 1999)

ICD-7	Site	Number of carcinomas		Incidence/10^5 person-years		RR[a] (95% CI) for non-EPIC compared with EPIC
		EPIC	Non-EPIC	EPIC	Non-EPIC	
151	Stomach	40	78	16.0	22.0	1.42 (0.97–2.08)
153	Colon (excluding rectum)	131	220	52.6	62.3	1.24 (1.00–1.54)
154	Rectum	110	141	44.1	39.9	0.93 (0.72–1.19)
153–154	Colon or rectum	236	358	95.0	102	1.11 (0.94–1.31)
157	Pancreas	40	84	16.0	23.7	1.53 (1.05–2.24)
162	Trachea, bronchus, lung or pleura	146	409	58.5	116	2.02 (1.67–2.44)
170[b]	Breast (invasive)	412	486	279	247	0.90 (0.79–1.02)
170[b]	Breast (cis + invasive)	475	530	324	270	0.84 (0.74–0.95)
171[b]	Cervix uteri (invasive)	22	61	14.5	30.4	2.07 (1.27–3.38)
171[b]	Cervix uteri (cis + invasive)	148	198	106	107	0.96 (0.77–1.19)
172[b]	Corpus uteri	69	83	45.7	41.3	0.94 (0.68–1.29)
175[b]	Ovary	48	71	31.7	35.3	1.14 (0.79–1.64)
177[b]	Prostate	354	396	366	263	0.79 (0.68–0.91)
180	Kidney	58	87	23.2	24.6	1.09 (0.78–1.52)
181	Urinary tract, (excluding kidney)	115	206	46.1	58.4	1.29 (1.03–1.62)
190	Malignant melanoma of skin	113	115	45.5	32.6	0.61 (0.48–0.79)
191	Skin (melanoma excl.)	91	96	36.5	27.2	0.77 (0.58–1.02)
193	Nervous system	67	97	26.8	27.5	1.05 (0.77–1.43)
200, 201	Malignant lymphoma	86	97	34.5	27.4	0.82 (0.62–1.10)

[a]Adjusted for sex and age at definition of original cohort, 1 January 1991. [b]Adjusted for age.

whether an autopsy had been performed or not was available). The same numbers for the non-EPIC cohort was 3515 deaths, 1478 during and 2037 following end of recruitment. Out of these, 2222 deaths were diagnosed prior to 31 December 1997. Among deaths that occurred prior to 31 December 1997, an

Table 4. Mortality in EPIC and non-EPIC cohorts

Period	Group	No. of individuals	No. of deaths	Mortality/10^5	RR[a] (95% CI)	Sex-specific RR[b] (95% CI)	
						Men	Women
1 January 1991 to	EPIC cohort	28 098	293	182	1.00 (ref.)	1.00	1.00
25 September 1996	Non-EPIC	40 807	1478	640	3.55 (3.13–4.03)	3.50 (2.98–4.11)	3.62 (2.97–4.42)
	Total	68 905	1771	452			
26 September 1996 to	EPIC cohort	27 759	689	771	1.00 (ref.)	1.00	1.00
31 December 1999	Non-EPIC	38 925	2037	1655	2.21 (2.03–2.41)	2.11 (1.88–2.36)	2.38 (2.08–2.73)
	Total	66 684	2726	1283			

[a]Adjusted for sex and age at definition of original cohort, 1 January 1991. [b]Adjusted for age.

Table 5. Cause-of-death in EPIC and non-EPIC cohorts during baseline examination until end of follow-up with regard to cause of death (1 January 1991 to 31 December 1997)

ICD-9	Site	Number of deaths		Mortality/10^5 person-years		RR[a] (95% CI) for non-EPIC compared with EPIC
		EPIC	Non-EPIC	EPIC	Non-EPIC	
140–239	All tumours	251	781	128	279	2.24 (1.95–2.59)
153–154	Colon or rectum cancer	24	77	12.3	27.5	2.34 (1.48–3.70)
162	Lung cancer	66	188	33.7	67.3	2.02 (1.53–2.68)
174[b]	Breast cancer	17	74	14.3	46.6	3.31 (1.95–5.61)
185[b]	Prostate cancer	19	38	24.7	31.4	1.39 (0.80–2.41)
390–459	All cardiovascular disease	191	767	97.6	274	2.85 (2.44–3.35)
410–414	Ischaemic heart disease	122	485	62.4	174	2.81 (2.30–3.42)
430–438	Stroke	33	113	16.9	40.4	2.49 (1.69–3.67)
800–999	Injuries and poison	25	139	12.8	49.7	3.81 (2.49–5.83)
	All deaths	527	2222	269	795	3.01 (2.74–3.31)

[a]Adjusted for sex and age at definition of original cohort, 1 January 1991.
[b]Adjusted for age.

autopsy had been performed in 48.6% of cases in the EPIC cohort and in 51.5% of deceased non-participants.

During baseline examination, mortality was more than three times higher in non-participants than in the EPIC cohort (Table 4). During the period following end of baseline examination, mortality was about twice as high in non-EPIC subjects as in participants (Table 4). Both willingness to participate and subsequent mortality may be associated with the same factors but it may also be that disease prior to end of baseline examination may have affected the ability to take part.

From beginning of recruitment until end of follow-up with regard to specific cause of death, mortality in the non-EPIC cohort was three times higher than in participants (Table 5). Death from cancer in non-EPIC subjects was 2.24 times more common, death from breast cancer 3.31 times higher and death from ischaemic heart disease 2.81 times more frequent than in participants. Furthermore, the non-EPIC cohort had a 3.81 times higher mortality from injuries and intoxication than had the EPIC cohort.

Discussion

This study has five main findings. First, the incidence of first time malignancy prior to recruitment was higher in participants as compared to non-participants. Secondly, the incidence of first time malignancy during recruitment was higher in non-participants. Thirdly, the mortality was higher in non-participants during recruitment. Fourthly, the mortality in non-participants stays higher than that in participants following end of baseline examination. And lastly, the comparison of questionnaire data from the EPIC cohort with a mailed health survey implies that the MDCS, with a participation rate of 38.9%, has the same socio-demographic structure and prevalence of smoking and obesity as a study with a 74.6% participation rate. However, several methodological issues have to be considered.

One such issue is related to the validity of cancer and mortality registries. The Swedish Cancer Registry is nation-wide and registration by both clinicians and pathologists is implemented by law. In Malmö, recent studies have reported a completeness of 99% and a correctness of 96% in regard of breast cancer diagnosed at Malmö University Hospital between 1961 and 1991 (Garne, 1996). Similar results have been reported from other parts of the country. The Cancer Registry was set up in 1958 and the oldest birth cohort in the MDCS was then 35 years of age. It is therefore possible that we have missed some tumours in early life but cancer incidence before age 35 is very low.

Even if completeness is high regarding diagnosed cancers, the probability of having a tumour detected may differ between groups. A delay between first symptoms and diagnosis (patient's or doctor's delay) or low participation in preventive health projects, such as mammography screening, may have the same determinants as participation in a cohort study like the MDCS (e.g. low health awareness and unwillingness to seek medical help). The finding that women in the EPIC cohort had a higher incidence of breast cancer may in fact be a result of high participation in mammographic screening instead of a truly increased risk of breast cancer.

The completeness of the Swedish Cause of Death Registry, as compared to the national population registry, was 99.64% in 1996 (National Board of Health and Welfare, 1998b). Some deaths may have been missed, but we have nothing to support the possibility that participants and non-participants would differ with regard to completeness of registration. The validity of information on specific cause-of-death is related to the autopsy frequency. This has traditionally been high in Malmö, close to 80%, but the rate has declined during the last decades (Garne, 1996). The autopsy rate is, however, kept high within the MDCS (Berglund *et al*., 1993). Up until 1997, the autopsy rate was similar in participants and non-participants and even slightly higher in the non-EPIC cohort. Therefore, the probability of correct diagnosis ought to have been similar in the two groups.

The comparison between the EPIC and the HSM:94 cohorts may have been biased as the questionnaire was distributed in different ways. Within the MDCS study, the questionnaire was reviewed together with the respondent at the second visit in order to complete missing values. This was obviously not possible in the mailed health survey. However, the only question in the HSM:94 that had a substantial number of missing values was that on alcohol consumption. It is well known that self-reported alcohol habits have low validity. However, it is reasonable to assume that this is mainly a problem concerning high consumers and that it may still be possible to compare the proportions of teetotallers in the two cohorts.

Weight (body mass index) was assessed by different methods in the MDCS and the HSM:94 cohorts:

direct measurement and self-reported information, respectively. However, self-reported height and weight have previously been shown to have high validity (Klag *et al*., 1993).

Our findings are in line with other studies that have found a higher overall mortality (Trell *et al*., 1985; Osler and Schroll, 1992) and a higher mortality from cancer (Trell *et al*., 1985; Rosengren *et al*., 1987; Walker *et al*., 1987) in non-participants than in participants. Participants and non-participants may differ in many other respects. The possibility that socio-economic status is associated with participation is supported in this study by high incidence in the EPIC cohort of breast cancer, prostate cancer, malign melanoma, cancers that are more common in affluent groups (Schottenfeld and Fraumeni, 1996; Lund Nilsen *et al*., 2000). Furthermore, the non-EPIC cohort had a high incidence of invasive cervix uteri cancer, a tumour associated with low socio-economic status (Schottenfeld and Fraumeni, 1996). Several cohort studies have indeed found a low participation rate in subjects with low educational level and poor social status (Bergstrand *et al*., 1983; Haglund *et al*., 1983; Rosengren *et al*., 1987; Persson *et al*., 1994; Jackson *et al*., 1996; Livingston *et al*., 1997).

This is not supported by the fact that educational level and the proportion of manual workers in the EPIC cohort were similar to the distribution in the HSM:94. However, our data show that the MDCS has the same socio-demographic structure as a study carried out in the same population with a participation rate of 74.6%. It is still possible that those who did and did not participate in the health survey differ in socio-economic status.

Lifestyle may also differ between participants and non-participants. This is suggested in this study by the fact that non-participants had a higher incidence of lifestyle-related cancers, such as pancreatic cancer (smoking, alcohol; Schottenfeld and Fraumeni, 1996), lung cancer (smoking; Schottenfeld and Fraumeni, 1996), and urinary tract tumours (smoking; Schottenfeld and Fraumeni, 1996). Non-participants also had a higher mortality from conditions associated with lifestyle factors, such as cardiovascular disease (Berglund *et al*., 1996), lung cancer (smoking; Schottenfeld and Fraumeni, 1996) and colorectal cancer (high dietary fat intake; Schottenfeld and Fraumeni, 1996). Some lifestyle factors have indeed been associated with low participation in other cohort studies; for example, smoking (Criqui *et al*., 1978; Persson *et al*., 1994; Jackson *et al*., 1996), high alcohol intake (Haglund *et al*., 1983; Rosengren *et al*., 1987; Persson *et al*., 1994), hypertension (Jackson *et al*., 1996; Bengtsson *et al*., 1997; Hill *et al*., 1997) and dietary habits (Van't Hof *et al*., 1991; Osler and Schroll, 1992). The comparison with the HSM:94 did not show any substantial differences in smoking habits and the proportion of teetotallers or subjects with obesity but the HSM:94 may have been subject to the same selection bias as the EPIC cohort.

The proportion of subjects born abroad was different in the MDCS and in the HSM:94. It is well known that factors such as tobacco consumption use of alcohol and diet differ between ethnical groups. The situation in Malmö is, however, complex as there are many immigrants from other Nordic countries as well as from outside Europe, and the association between different lifestyle habits and ethnic origin in Malmö remains to be evaluated in future studies.

Lifestyle factors and socio-economic class may be associated with both willingness to participate and subsequent risk of developing disease, but apart from a personal decision to participate, some conditions (e.g. advanced disease) will make it difficult to take part in the baseline examination. This is indicated in this study from the overall higher mortality, specifically from tumours and cardiovascular disease that was seen in non-participants. Increased morbidity may indeed be the very reason why they did not participate. This is also in line with the finding that the EPIC cohort had a lower proportion of subjects who considered themselves to have poor health than had the HSM:94. Following baseline examination, participants who develop disease will still be participants, and this will result in less pronounced differences in mortality between this group and non-participants. The rapid increase in mortality rates among participants following end of recruitment may thus be explained.

We conclude that mortality is higher in non-participants than in participants during recruitment and follow-up. It is also suggested that non-participants may have a lower cancer incidence prior to recruitment but a higher incidence during the recruitment period. This may have implications for the interpretation of future studies within the MDCS as cancer incidence, mortality and exposure rates may not be directly applicable to the general population. Furthermore, data analysis may be affected if there is a selection of subjects among whom there is a low prevalence of a potential risk factor (e.g. the proportion of smokers), or a very low incidence of a certain

end point (e.g. bladder cancer). A lack of statistically significant associations may thus be a reflection of poor statistical power.

However, in future studies of associations between exposure and disease, the main topic will be whether there is a big enough difference in exposure levels within the cohort (e.g. intake of dietary fat). Even if the EPIC cohort is selected towards better health, and possibly higher socio-economic status and healthier lifestyle, it will still be possible to study associations and causal pathways as long as there is a heterogeneous distribution of the studied factor.

Acknowledgements—The authors wish to thank Ms Ulla Lundgren, administrative assistant whose help was crucial in the definition of the cohort and who kindly answered all questions about the routines at the baseline examination. Financial support has been given by the Swedish Cancer Foundation, The Swedish Medical Research Council, The European Commission, the City of Malmö, the Swedish Dairy Association and the Albert Påhlsson Foundation.

References

Bengtsson C, Ahlqwist M, Andersson K, *et al*. (1997). The prospective population study of women in Gothenburg, Sweden, 1968–69 to 1992–93. A 24-year follow-up study with special reference to participation, representativeness, and mortality. *Scand J Prim Health Care* **15:** 214–19.

Berglund G (1993). The Malmö Diet and Cancer Study. Design biological bank and biomarker programme. *J Intern Med* **233:** 39–40.

Berglund G, Elmståhl S, Janzon L, Larsson SA (1993). The Malmö Diet and Cancer Study—Design and Feasibility. *J Intern Med* **233:** 45–51.

Berglund G, Eriksson K F, Israelsson B, *et al*. (1996). Cardiovascular risk group and mortality in an urban Swedish male population: The Malmö Preventive Project. *J Intern Med* **239:** 489–97.

Bergstrand R, Vedin A, Wilhelmsson C, Wilhelmsen L (1983). Bias due to non-participation and heterogenous sub-groups in population surveys. *J Chronic Dis* **36:** 725–8.

Criqui MH, Barrett-Connor E, Austin M (1978). Differences between respondents and non-respondents in a population-based cardiovascular disease study. *Am J Epidemiol* **108:** 367–72.

Elmståhl S, Gullberg B, Riboli E, Saracci R, Lindgärde F (1996). The Malmö Food Study: the reproducibility of a novel diet history method and an extensive food frequency questionnaire. *Eur J Clin Nutr* **50:** 134–42.

EPIC (1998). *Guidelines for Collection of End-Point Data in the EPIC Study*. International Agency for Research on Cancer, Lyon, France.

Garne JP (1996). Invasive breast cancer in Malmö 1961–1992 – an epidemiological study. Thesis, Lund University, Malmö, Sweden.

Haglund BJ, Isacsson SO, Ryden L, Råstam L (1983). Health profile of Skaraborg County 1977. A Swedish rural cross-sectional study. *Scand J Prim Health Care* **1:** 102–13.

Hanson BS, Östergren PO, Elmståhl S, Isacsson SO, Ranstam J (1997). Reliability and validity assessments of measures of social networks, social support and control – results from The Malmö Shoulder and Neck Study. *Scand J Soc Med* **25:** 249–57.

Hill A, Roberts J, Ewings P, Gunnell D (1997). Non-response bias in a lifestyle survey. *J Public Health Med* **19:** 203–7.

Jackson R, Chambless LE, Yang K, *et al*. (1996). Differences between respondents and nonrespondents in a multicenter community-based study vary by gender ethnicity. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *J Clin Epidemiol* **49:** 1441–6.

Klag MJ, He J, Mead LA, *et al*. (1993). Validity of physicians' self-reports of cardiovascular disease risk factors. *Ann Epidemiol* **3:** 442–7.

Lindström M, Bexell A, Hanson B S, Isacsson S-O (1995). *The Health Situation in Malmö: Report from a Mailed Questionnaire Survey, Spring 1994* (in Swedish). Department of Community Medicine, Malmö University Hospital, Malmö, Sweden.

Livingston PM, Lee SE, McCarty CA, Taylor HR (1997). A comparison of participants with non-participants in a population-based epidemiologic study: the Melbourne Visual Impairment Project. *Ophthalmic Epidemiol* **4:** 73–81.

Lund Nilsen TI, Johnsen R, Vatten LJ (2000). Socio-economic and lifestyle factors associated with the risk of prostate cancer. *Br J Cancer* **82:** 1358–63.

National Board of Health and Welfare (1998a). *Cancer Incidence in Sweden 1996*. The National Board of Health and Welfare, Stockholm.

National Board of Health and Welfare (1998b). *Causes of Death 1996*. The National Board of Health and Welfare, Stockholm.

Osler M, Schroll M (1992). Differences between participants and non-participants in a population study on nutrition and health in the elderly. *Eur J Clin Nutr* **46:** 289–95.

Pero RW, Olsson A, Bryngelsson C, *et al*. (1998a). Quality control program for storage of biologically banked blood specimens in the Malmö Diet and Cancer Study. *Cancer Epidemiol Biomarkers Prev* **7:** 803–8.

Pero RW, Olsson A, Bryngelsson C, *et al*. (1998b). Feasibility and quality of biological banking of human normal and tumor tissue specimens as sources of DNA for the Malmö Diet and Cancer Study. *Cancer Epidemiol Biomarkers Prev* **7:** 809–12.

Persson J, Hedblad B, Israelsson B, Berglund G (1997). Factors associated with ultrasound determined atherosclerosis in carotid artery. In *Ultrasound and Atherosclerosis*, (Persson J); Thesis, Lund University, Malmö, Sweden.

Persson LG, Lindström K, Lingfors H, Bengtsson C (1994). A study of men aged 33–42 in Habo, Sweden with special reference to cardiovascular risk factors. Design, health profile and characteristics of participants and non-participants. *Scand J Soc Med* **22:** 264–72.

Riboli E (1992). Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol* **3:** 783–91.

Riboli E, Kaaks R (1997). The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26(Suppl 1):** S6–14.

Riboli E, Elmståhl S, Saracci R, Gullberg B, Lindgärde F (1997). The Malmö Food Study: validity of two dietary assessment methods for measuring nutrient intake. *Int J Epidemiol* **26(Suppl 1):** S161–73.

Rosengren A, Wilhelmsen L, Berglund G, Elmfeldt D (1987). Non-participants in a general population study of men, with special reference to social and alcoholic problems. *Acta Med Scand* **221:** 243–51.

Schottenfeld D, Fraumeni J F (1996). *Cancer Epidemiology and Prevention*, 2nd edn. Oxford University Press, New York.

Southern Swedish Regional Tumour Registry (1994). *Cancer Incidence in Southern Sweden 1988–1992*. The Southern Swedish Regional Tumour Registry, Lund.

Trell E, Kristenson H, Petersson B (1985). A risk factor approach to the alcohol-related diseases. *Alcohol Alcohol* **20:** 333–45.

Walker M, Shaper AG, Cook DG (1987). Non-participation and mortality in a prospective study of cardiovascular disease. *J Epidemiol Commun Health* **41:** 295–9.

Van't Hof MA, Hautvast JG, Schroll M, Vlachonikolis IG (1991). Design, methods and participation. Euronut SENECA investigators. *Eur J Clin Nutr* **45(Suppl 3):** 5–22.