

Research article

Open Access

## The Cornell Health Interview Survey Follow-Up (CHIS.FU) Study: design, methods, and response rate

Montse Garcia<sup>1,2</sup>, Anna Schiaffino<sup>1</sup>, Esteve Fernandez\*<sup>1,3</sup>, Merce Marti<sup>4</sup>, Esteve Salto<sup>3,5</sup>, Gloria Perez<sup>5</sup>, Merce Peris<sup>1</sup>, Carme Borrell<sup>6</sup>, F Javier Nieto<sup>7</sup> and Josep Maria Borrás<sup>1,3</sup>

Address: <sup>1</sup>Cancer Prevention and Control Unit, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Spain, <sup>2</sup>Department of Methodology, University of Barcelona, Spain, <sup>3</sup>Department of Public Health, University of Barcelona, Spain, <sup>4</sup>Department of Public Health, Cornell de Llobregat City Council, Spain, <sup>5</sup>Ministry of Health, Autonomous Government of Catalonia, Barcelona, Spain, <sup>6</sup>Municipal Institute of Public Health, Barcelona, Spain and <sup>7</sup>Department of Population Health Sciences, University of Wisconsin, United States

Email: Montse Garcia - mgarciam@ico.scs.es; Anna Schiaffino - aschiaffino@ico.scs.es; Esteve Fernandez\* - efernandez@ico.scs.es; Merce Marti - mmarti@aj-cornella.es; Esteve Salto - esalto@dsss.scs.es; Gloria Perez - gperez@dsss.scs.es; Merce Peris - m.peris@ico.scs.es; Carme Borrell - cborrell@imsb.bcn.es; F Javier Nieto - fjnieto@facstaff.wisc.edu; Josep Maria Borrás - jmborras@ico.scs.es

\* Corresponding author

Published: 7 March 2003

Received: 18 November 2002

BMC Public Health 2003, 3:12

Accepted: 7 March 2003

This article is available from: <http://www.biomedcentral.com/1471-2458/3/12>

© 2003 Garcia et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** The aim of this report is to describe the main characteristics of the design, including response rates, of the Cornell Health Interview Survey Follow-up Study.

**Methods:** The original cohort consisted of 2,500 subjects (1,263 women and 1,237 men) interviewed as part of the 1994 Cornell Health Interview Study. A record linkage to update the address and vital status of the cohort members was carried out using, first a deterministic method, and secondly a probabilistic one, based on each subject's first name and surnames. Subsequently, we attempted to locate the cohort members to conduct the phone follow-up interviews. A pilot study was carried out to test the overall feasibility and to modify some procedures before the field work began.

**Results:** After record linkage, 2,468 (98.7%) subjects were successfully traced. Of these, 91 (3.6%) were deceased, 259 (10.3%) had moved to other towns, and 50 (2.0%) had neither renewed their last municipal census documents nor declared having moved. After using different strategies to track and to retain cohort members, we traced 92% of the CHIS participants. From them, 1,605 subjects answered the follow-up questionnaire.

**Conclusion:** The computerized record linkage maximized the success of the follow-up that was carried out 7 years after the baseline interview. The pilot study was useful to increase the efficiency in tracing and interviewing the respondents.

### Background

Follow-up for mortality and incidence of cancer and cardiovascular disease of a population-based cohort can pro-

vide useful information from an epidemiological and public health perspective as well as for health services organization. Population-based cohort studies are however

still scarce and most of the currently available information is based on general population cohorts followed up in the United States [1–5]. In Spain, there is not a well established tradition of follow-up studies of representative population-based cohorts. A follow-up of a sample of the elderly people in Barcelona has provided useful information about mortality determinants [6]. In Spain, there are cohorts of volunteers or workers conducted as part of integrated European projects, such as the cardiovascular risk factors MONICA project [7] and the EPIC study [8,9] on diet and cancer.

We planned to establish a representative cohort of the Cornellà population based on a previously conducted survey of 2,500 people, the 1994 Cornellà Health Interview Survey. The specific aims of the Cornellà Health Interview Survey Follow Up (CHIS.FU) were: a) to analyze changes in smoking habit, alcohol consumption, and level of physical activity between 1994 (baseline interview) and 2002 (follow-up interview), and their sociodemographic determinants; b) to determine the mortality and incidence of cardiovascular diseases and cancer in the cohort and to analyze their association with socioeconomic status, self-perceived health, life styles, and chronic conditions. Secondary to this objective, the CHIS.FU will provide new cross-sectional information to be used in further follow-up, otherwise we could use that data for the following analyses: c) to describe risk perception beliefs on cancer; d) to analyze social support variables and their association with self-perceived health, life-styles, and use of health care services; and, e) to analyze health locus of control in relation to lifestyles and mortality.

It is well known that the success of any survey depends on achieving a high response rate. Numerous factors can influence the participation rate, including the methods used to contact subjects, the effort required from the participants, characteristics of the target population, and its interest in the research [10,11]. Efforts to increase the response rate have been classified by timing (preliminary, concurrent and follow up efforts) and by technique (questionnaire length, size, survey sponsorship, trained interviewers) [12].

Non-response reduces the effective sample size and can introduce bias [13] and it is the major concern for survey research, in particular, cohort studies [14]. Three types of attrition (failure to locate, refusal to participate, mortality) have been described [15] but no matter how careful the researchers are in implementing tracing procedures and in keeping subjects motivated to continue to take part in the study, there will always be such losses [16,17].

Losses in the follow-up do not necessarily invalidate the research as such [18,19]. Researchers, however, routinely

use strategies to minimize losses to follow up and consider whether such losses bias or not the results obtained from the study [20]. Ascertainment of full and accurate information at the beginning of the study is extremely important to avoid failure to locate cohort members in future contacts [17].

This paper focuses on the rationale, design, methods, and description of the response rate of the CHIS.FU study, with special emphasis on efforts to update the address and vital status of the participants that took place before the follow-up interview as well as a description of the different techniques that were used to improve the response rate.

## Methods

### *Inception cohort*

The baseline health and sociodemographic characteristics of subjects reported in this paper were obtained from a cross-sectional survey, the Cornellà Health Interview Survey (CHIS) carried out in 1994 [21–23]. Cornellà de Llobregat <http://www.cornellaweb.com> is a town of approximately 85,000 inhabitants, mainly working- and middle class, located on the metropolitan area of Barcelona, Catalonia (Spain). A representative sample of the non-institutionalized population (all ages) of 2,500 people (1,263 women and 1,237 men) was selected by simple random sampling from the Census and interviewed face-to-face, during 1994 (12 months) to avoid seasonal variations.

The variables studied in the CHIS included sociodemographic and personal information (place of birth, age, sex, educational level and social class); height and weight, health behaviors (smoking habit, alcohol consumption, and physical activity); chronic conditions; self-perceived health; and use of health care services. Detailed information of the survey is available elsewhere [21].

The inclusion of participants in the cohort was based on the interview date, starting in 1994 with follow-up until death, migration or censoring date (30/07/02).

### *Record linkage*

In April 2000, we implemented a computerized record linkage to update the address and vital status of all 2,500 participants, with the objective of optimizing the response rate and improving direct tracing of cohort members. Afterwards, we attempted to trace the cohort members to conduct the phone follow-up interviews.

The data collected in 1994 were stored according to current procedures to ensure confidentiality. An independent computerized file with the identification of data (name, surnames, address, phone number, and date of birth) was

**Table 1: Response rate (%) and tracking results according to sociodemographic characteristics and health related variables. The Cornellia Health Interview Survey Follow-up Study (CHIS.FU).**

| Baseline Characteristics      |                   | Participants |                      |                  | Non-participants  |                           |              |                    |                    |
|-------------------------------|-------------------|--------------|----------------------|------------------|-------------------|---------------------------|--------------|--------------------|--------------------|
|                               |                   | n            | General Q. n = 1,438 | Proxy Q. n = 170 | Refusal           |                           | Dead n = 147 | Emigration n = 425 | Non-traced N = 197 |
|                               |                   |              |                      |                  | Refusal Q. n = 94 | Refusal without Q. n = 29 |              |                    |                    |
| Sex                           | Males             | 1,236        | 668 (54.0)           | 85 (6.9)         | 42 (3.4)          | 22 (1.8)                  | 88 (7.1)     | 210 (17.0)         | 122 (9.8)          |
|                               | Females           | 1,262        | 770 (61.0)           | 85 (6.7)         | 52 (4.1)          | 7 (0.6)                   | 59 (4.7)     | 215 (17.0)         | 75 (5.9)           |
| Age in 1994                   | 0-14              | 354          | 153 (43.2) *         | 114 (32.2)       | 4 (1.1)           | 6 (1.7)                   | 2 (0.3)      | 53 (15.0)          | 23 (6.5)           |
|                               | 15-44             | 1,078        | 633 (58.6)           | 5 (0.5)          | 32 (3.0)          | 13 (1.2)                  | 6 (0.6)      | 279 (25.8)         | 111 (10.3)         |
|                               | 45-64             | 692          | 506 (73.1)           | 14 (2.0)         | 34 (4.9)          | 7 (1.1)                   | 30 (4.4)     | 63 (9.2)           | 36 (5.3)           |
|                               | ≥ 65              | 376          | 146 (38.8)           | 37 (9.8)         | 24 (6.4)          | 3 (0.8)                   | 109 (29.0)   | 30 (8.0)           | 27 (7.2)           |
| Place of birth                | Cornellia         | 979          | 524 (53.5)           | 108 (11.0)       | 20 (2.0)          | 13 (1.3)                  | 10 (1.0)     | 217 (22.2)         | 86 (8.8)           |
|                               | Rest of Catalonia | 286          | 170 (59.4)           | 9 (3.1)          | 7 (2.4)           | 1 (0.3)                   | 24 (8.3)     | 54 (18.8)          | 23 (8.0)           |
|                               | Rest of Spain     | 1,202        | 730 (60.7)           | 51 (4.2)         | 67 (5.6)          | 14 (1.2)                  | 112 (9.3)    | 147 (12.2)         | 79 (6.6)           |
|                               | Abroad            | 33           | 14 (42.4)            | 2 (6.1)          | -                 | 1 (3.0)                   | 1 (3.0)      | 7 (21.2)           | 9 (27.3)           |
| Educational level in 1994     | <10 years old     | 205          | 40 (19.5) **         | 113 (55.1)       | 3 (1.5)           | 4 (2.0)                   | 1 (0.5)      | 37 (18.0)          | 7 (3.4)            |
|                               | Less than primary | 423          | 217 (51.3)           | 36 (8.5)         | 31 (7.3)          | 4 (0.9)                   | 70 (16.5)    | 40 (9.5)           | 23 (5.4)           |
|                               | Primary           | 1,309        | 857 (65.3)           | 17 (1.3)         | 51 (3.9)          | 17 (1.3)                  | 70 (5.4)     | 185 (14.1)         | 114 (8.7)          |
|                               | Secondary         | 470          | 273 (58.1)           | 1 (0.2)          | 7 (1.5)           | 4 (0.9)                   | 4 (0.9)      | 137 (29.2)         | 43 (9.1)           |
|                               | University        | 86           | 50 (58.1)            | 1 (1.2)          | 2 (2.3)           | -                         | -            | 24 (27.9)          | 10 (10.5)          |
| Self-perceived health in 1994 | Good              | 1694         | 981 (57.9)           | 122 (7.2)        | 57 (3.3)          | 23 (1.4)                  | 40 (2.4)     | 331 (19.5)         | 142 (8.3)          |
|                               | Not bad           | 648          | 386 (59.6)           | 37 (5.7)         | 27 (4.1)          | 4 (0.6)                   | 70 (10.8)    | 78 (12.0)          | 47 (7.2)           |
|                               | Bad               | 156          | 71 (45.6)            | 11 (7.1)         | 10 (6.5)          | 2 (1.4)                   | 37 (23.8)    | 16 (10.3)          | 8 (5.2)            |
| Smoking status in 1994        | Smoker            | 608          | 356 (58.6)           | 5 (0.8)          | 23 (3.8)          | 11 (1.8)                  | 22 (3.6)     | 130 (21.4)         | 60 (9.9)           |
|                               | Former smoker     | 211          | 119 (56.3)           | 7 (3.3)          | 5 (2.4)           | 1 (0.5)                   | 27 (12.8)    | 36 (17.1)          | 16 (7.6)           |
|                               | Never smoker      | 1255         | 793 (63.1)           | 30 (2.4)         | 60 (4.7)          | 11 (0.9)                  | 71 (5.6)     | 198 (15.8)         | 94 (7.5)           |
| Alcohol consumption in 1994   | Non drinker       | 447          | 284 (63.4)           | 12 (2.7)         | 15 (3.3)          | 4 (0.9)                   | 27 (6.0)     | 73 (16.3)          | 33 (7.4)           |
|                               | Moderate drinker  | 698          | 433 (62.1)           | 9 (1.3)          | 29 (4.2)          | 9 (1.3)                   | 23 (3.3)     | 136 (19.5)         | 58 (8.3)           |
|                               | Risk drinker      | 133          | 77 (57.9)            | 1 (0.8)          | 6 (4.5)           | 1 (0.7)                   | 6 (4.5)      | 30 (22.6)          | 12 (9.0)           |
| Cronic conditions in 1994     | 0                 | 1117         | 624 (55.8)           | 107 (9.6)        | 30 (2.7)          | 13 (1.1)                  | 19 (1.7)     | 233 (20.8)         | 93 (8.3)           |
|                               | 1                 | 530          | 313 (59.1)           | 20 (3.8)         | 22 (4.2)          | 8 (1.5)                   | 21 (3.9)     | 96 (18.1)          | 50 (9.4)           |
|                               | 2                 | 308          | 184 (59.7)           | 10 (3.2)         | 13 (4.2)          | 3 (1.0)                   | 30 (9.7)     | 48 (15.6)          | 20 (6.5)           |
|                               | ≥ 3               | 541          | 317 (58.6)           | 32 (5.9)         | 29 (5.4)          | 5 (0.9)                   | 77 (14.2)    | 47 (8.7)           | 34 (6.3)           |

\* These participants were <15 years old in 1994 but ≥ 15 years in 2002, and hence answered the general questionnaire. \*\*These participants were <10 years old in 1994 but ≥ 15 years in 2002, and hence answered the general questionnaire.

created for linkage with the *Local Census of Cornellia* in order to update participants' data (vital status, address). The record linkage was carried out in 2 steps. First, a deterministic method based on the first name and surnames (in Spain, father and mother family names are retained by descendants) of each participant. By this method, the matching took place when there was a complete correspondence between all selected variables (name, first surname and second surname). Second, a probabilistic method based on the partial correspondence of the same variables was used [24]. The record linkage is difficult because of composite first and family names and person's names including "ñ" or other characters (eg.ç) or accented vowels [25]. Thus, a person names Juan Manuel Morales Gracia can be registered as J. Manuel Morales or Juan M. Morales Gracia or Juan Morales Gracia and if the record linkage is based

on a probabilistic method the percentage of success matching is much higher.

All these procedures were carried out by authorized personnel from the City Council of Cornellia, following confidentiality rules.

**Questionnaires and pilot study**

We designed three types of questionnaires to obtain the follow-up information by telephone interview: 1) a general questionnaire for those participants aged ≥ 15 years who directly could respond the interview; 2) a proxy questionnaire: a relative was asked to answer the interview if the participants were less than 15 years old and/or were disabled to respond by themselves; and 3) a short refusal questionnaire for those people who did not agree to be in-

interviewed: they were asked to respond three questions about self-perceived health, smoking status, and educational level. Thus, if full participation was not feasible, at least a few potential explanatory variables could be collected with their consent [19]. Questionnaires were available in Catalan and Spanish, the two official languages in Catalonia, Spain. All subjects who agreed to answer the follow-up interview gave their oral informed consent. We also asked consent for future follow-up interviews and those who agreed to be re-interviewed in following interviews will be contacted in some years.

The questionnaires were pre-tested in *ad hoc* interviews to volunteers to check its comprehension, time length, and overall feasibility. After that, a pilot study was carried out among a random sample of 100 participants (selected after the record linkage) from November to December 2001. The pilot study enabled us to consider some factors related to participation before the start of the field work, such as time cost, response rates, data quality, and acceptability to subjects.

The response rate obtained in the pilot study ( $n = 100$ ) was 78.0%, 69% answered the general questionnaire and 9% the proxy questionnaire because they were minors and/or were disabled (table 1). The refusal rate was 5%, but almost 90% of non-responders (4% of sample of the pilot study) agreed to answer the short questionnaire (self-perceived health, smoking habit and educational level). From the total sample of the pilot study 9% of emigrations and 2% deceased persons were found. Finally, we could not trace 6% of the sample in the pilot study.

#### **Procedures for data collection**

A protocol was established with detailed instructions about how to carry various aspects of the survey (the introduction of the interviewer, the wording of the instructions, etc). After the pilot study the protocol was slightly modified as follows:

An informative letter signed by the Public Health town councilor and the principal investigator of the project was sent to the participants in the 1994 CHIS, in advance to the phone contact, describing the general purpose of the study and eliciting cooperation. In addition, a leaflet with the main results from the 1994 CHIS was attached. The letters were mailed weekly in waves of 120. A week after the mailing the cohort members were telephoned to ask their collaboration and answer the questionnaire.

Four shifts were established in order to trace the subjects and to conduct the interview: 1<sup>st</sup> shift from 10:00 to 13:00, 2<sup>nd</sup> shift from 12:00 to 14:00, and after lunch time from 15:00 to 16:00, 3<sup>rd</sup> shift from 16:00 to 19:00, and finally the 4<sup>th</sup> shift from 19:00 to 22:00. An extra shift for

those non-located in the previous shifts was established during week-ends. To obtain a complete telephone interview a maximum of 15 calls were made. Attempts to reach the participants were made in all shifts and during week-ends and again after several weeks. The results during phone tracking were written down.

In order to make the phone call, interviewers dialed the number and held the line for 7 full tones. A record of the result was left after each call. After 3 calls in a same shift (even being different days and hours) if it had not been possible to contact the individual or a relative, the questionnaire was passed to another phone shift. If the phone number was incorrect the entire process started again. First, we checked mistakes when dialing. We also checked if we had another telephone number and if not, the telephone was searched through white pages, telephone assistance, and the Local Census. Finally, a questionnaire was sent by mail to those cohort members who could not be reached by telephone.

Following the interview, an acknowledgement letter was sent to all the participants who responded the questionnaire to promote interest in the study and facilitate future participation. Also, as a reinforcement for further follow-up, a greeting card for the New Year was sent during the month of December, including a leaflet with the preliminary results of the follow-up.

To minimize future losses to follow-up we included an item asking for the name and phone number of one friend or relative not living with the participant who could be contacted if the cohort member is not traced.

Interviews were conducted by 4 interviewers subject to standardized training consisting of an exhaustive explanation on how to conduct the interview and simulation of the interview by means of role-playing techniques. The purpose of this training was, besides achieving the highest possible accuracy in responses and homogeneity in the administration of the questionnaire, to minimize refusals. Most refusals occur during the introduction in the first moment of the telephone call, so a personalized and carefully worded presentation was emphasized. A manual with step-by-step instructions for completion of the questionnaire was also used by the interviewers. Furthermore, they were showed how to edit the data as they were collected.

Data in the CHIS.FU were managed centrally at the Prevention and Cancer Control Unit from the Catalan Institute of Oncology, following the confidentiality rules for this type of data.

### Variables and statistical procedures

For the purposes of this report, we have analyzed the response rate (percentage of responses) in the pilot study and the overall cohort according to the three types of questionnaires. We have also taking into account the percentage of refusals, emigration, deaths and non-located subjects. We analyzed the response rate according to selected sociodemographic baseline characteristics, such as sex, age in 1994 (0–14, 15–44, 45–64, and  $\geq 65$  years), place of birth (the city of Cornellà, rest of Catalonia, rest of Spain, abroad), and educational level in 1994 (less than primary, primary, secondary, university, and we defined a special category for children aged  $<10$  years old) as well as some health related variable such as self-perceived health (good, poor, bad), smoking status (current smoker, former smoker and never), alcohol consumption (no consumption, 0 g/day; moderate consumption,  $\leq 40$  g/day in males and  $\leq 24$  g/day in females; and risk consumption,  $> 40$  g/day in males and  $> 24$  g/day in females) and chronic conditions, from a list of 16 chronic conditions usually included in health surveys were recoded as follows: 0 chronic conditions, 1 chronic condition, 2 chronic conditions and  $\geq 3$  chronic conditions.

## Results

### Record linkage

After the initial record linkage, almost 75% of subjects interviewed in 1994 were traced. Of these, 2% had died, 6% migrated from Cornellà and 5% had neither renewed their last municipal census documents nor announced they had moved. Furthermore, 195 (12.9%) had changed their address within Cornellà. Nevertheless, 130 differences according to date of birth were found, due to incorrect matches or to inaccurate information of file registers (initial error rate: 7%). Most of these inconsistent matches were found among residents in Cornellà ( $n = 90$ ), followed by those participants who had not updated their last municipal census information ( $n = 25$ ).

However, after using the probabilistic method the record linkage increased the proportion of subjects traced to 2,468 (98.7%) (figure 1). Of these, 91 (3.6%) were deceased, 259 (10.3%) had moved to other towns, and 50 (2.0%) had neither renewed their last municipal census data nor declared that they had moved, and 267 (10.7%) changed their address within the city.

### Field Work

In contrast to the high response rates in the pilot study, the response rate in the total cohort was 64.3%. 1,438 subjects (57.5%) answered directly to the questionnaire and 170 (6.8%) used the proxy questionnaire (table 1). We received 25 questionnaires by mail from the subjects that could not be contacted by telephone.

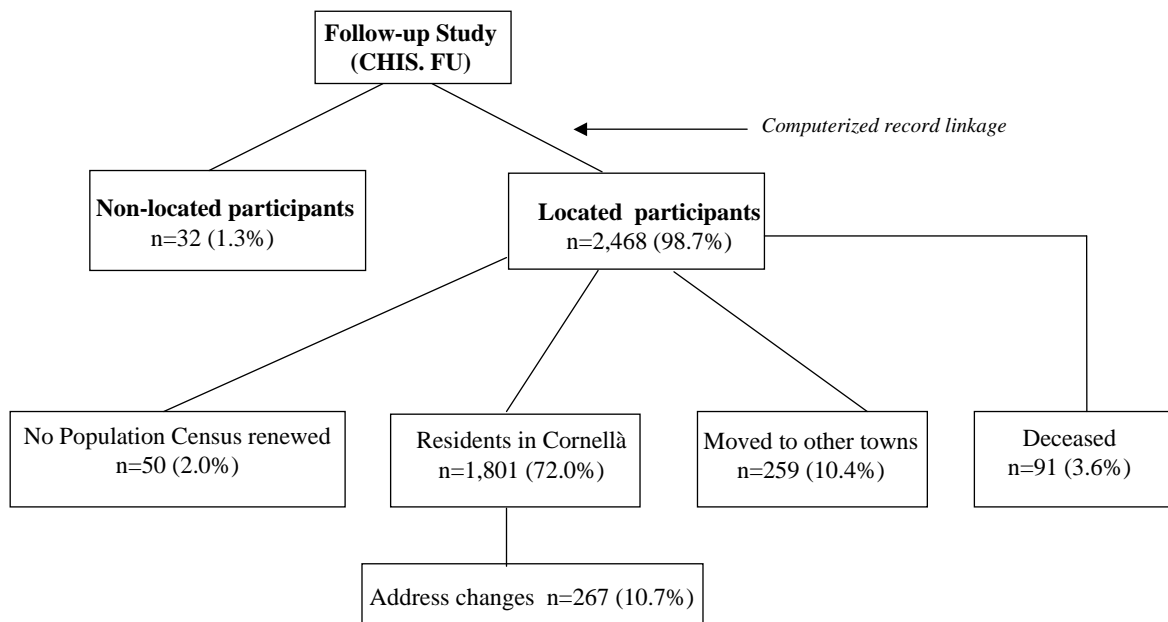
As in the pilot study, there was a 5% of refusal, and 3.8% of the overall cohort responded to the short questionnaire. We found that 17.0% of the cohort members had emigrated and 5.8% had died. For this 22.8% of the sample, we collected the date of censoring (data of migration or of death) from the Local Census or proxy respondents. The rate of traced subjects was therefore 92.1% ( $n = 2,303$ ). In 105 out of the 197 cohort members who were not traced, the telephone number in our files was incorrect and we were not able to update it.

Table 1 shows the response rate and tracking according to sociodemographic characteristics. Response rate was higher in women (67.7%) than in men (60.9%,  $p < 0.05$ ). Absolute refusal rate (no answer to the brief non-response questionnaire) was higher among men ( $n = 22$ , 1.8%). Only 7 women refused to respond the brief questionnaire (Table 1). The mortality rate was almost twice in men (7.1%) than in women (4.7%). We did not find gender differences in rates of out-migration (not shown). Finally, more men than women could not be traced (9.9% vs. 5.9% respectively,  $p < 0.05$ ). Participants aged 45–64 years answered more frequently the general questionnaire (73.1%) and the proxy questionnaire was mainly responded by a relative when cohort members were less than 15 years old or  $\geq 65$  years old. As expected, mortality was higher among cohort members aged  $\geq 65$  (29.0%). In relation to emigration, most of the cohort members who have moved to another town were aged 15–44 years.

The response rate among cohort members born out of Spain was lower (48.5%) than among those born in Spain (about 65%), although the number of foreigners was small ( $n = 33$ ).

According to the educational level, the highest response rate was obtained among those participants who had primary studies in 1994 (65.3%). The highest mortality rate was also obtained among the cohort members with less than primary studies (16.5%). Out-migrations were more frequent among cohort members with secondary (29.2%) and university (27.9%) education. The proportion of non-traced cohort members was higher among those with higher level of education (table 1).

The response rate and tracking according to selected baseline sociodemographic characteristics among men and women has also obtained (data not shown). The only relevant differences we have found have to do with non-traced cohort members. From those, most of men were aged 15–44 years (14.2%) and most of women were aged  $\geq 65$  years (11.2%). The response rate and tracking according to health related variables was also considered. People deceased within 1994–2002, declared a worse self-perceived health (25.2%) in the baseline interview and half



**Figure 1**  
Distribution of traced and non-traced cohort members of the Cornellà Health Interview Survey Follow-up (CHIS.FU) Study after record linkage.

of those people reported  $\geq 3$  chronic conditions (52.4%). In relation to life-style variables, non-traced subjects, refusers who did not answered the brief questionnaire and emigrated subjects where more frequently ever smokers. Subjects who emigrated and non-traced subjects were more likely to report a high consumption of alcohol in 1994 (7.0% vs. 6.1%) (Table 1).

**Discussion**

A preliminary effort to achieve an acceptable response rate in the CHIS.FU was to update the current address and vital status of the 2,500 cohort members. The record linkage has allowed to assess the vital status and to update contact information on cohort members. The results of a relatively simple linkage method confirm the successful tracing of almost 75% of the initial cohort, implying low cost and high efficiency.

The initial linkage was optimized applying a probabilistic method in the 670 cases that were not initially located as well as in the 130 matching errors, increasing the proportion of traced subjects from 73.2% to 98.7%. The main limitations to the interpretation of record linkage are duplication of identifiers and errors in creating or transmitting records [26]. So it is necessary the use of probabilistic techniques as well as a later manual revision to maximize its sensitivity and specificity [24].

Linkage involves by definition personal identifiers (name and surnames) and, thus, confidentiality issues need to be considered. Once the participants are located and have passed a second interview, no personal identification appears in the study database. The large societal benefits from cohort studies as well as the very small risks involved

in the individual participation in such studies must be emphasized [26].

The effort spent in locating cohort members improves the precision as well as the validity of the study results [17]. In this study we obtained a full response from 64.2% of cohort members. In our study, the cohort population is defined by geographic boundaries so the follow-up was made among those subjects who were alive and who were still living in Cornellà. Therefore, given that migrations and deaths are natural losses to follow-up [18], the actual participation rate after follow-up was 83.4% (1,608 interviews over 1,928 non-migrated and non-deceased cohort members). Although 94 cohort members refused to answer the general questionnaire, they provided basic information through the short questionnaire (total participation rate of 88.3%). Some authors have recommended to report different response rates according to the eligibility of subjects and results of follow-up [27].

These rates of participation in an epidemiological study are very high in our country [28], where there is scarce tradition to participate in epidemiological studies. Response rates may vary according to the interview method used [29]. We chose the telephone interview instead of mailed questionnaires given the socioeconomic status of the overall cohort members [30], and instead of the face-to-face interview because of logistic procedures and cost [31].

Activities to locate lost or hard-to-find participants have been continued until the participant's location and/or vital status have been ascertained, or until search strategies have been exhausted [12]. Findings are representative of the population only if those people who do not respond to the questionnaire do not differ in significant ways from those who do respond. If they do differ, the kind and degree of such differences must be carefully estimated so that the findings may be properly weighted to reflect more accurately the population under study [12].

It was also necessary and useful to test the questionnaire that has been designed for phone administration, to determine whether modifications were needed and lately to interview all the participants in the CHIS.FU Study. The point of pilot testing of procedures is to anticipate and eliminate as many problems as possible once the study is launched [32]. The pilot study had consequences in the later organization and in establishing the final protocol for the study implementation in order to increase the efficiency in the location and interview of the subjects. However, the pilot study presented a somehow optimistic view over the response rate than the final field work results. This could be attributed to random variation resulting from the limited sample size.

We have increased the response rate by using several strategies shown to be effective [13]. Minimizing losses to follow-up or attrition is important for two reasons: 1) a high response rate increases the power of the statistical analyses, and 2) attrition is likely to be non-random, it may also produce bias in the findings [19,33–35]. So we have to check if the characteristics of follow-up participants and drop-outs differ, and in that case, we have to estimate more accurately if those differences will bias the findings. Representativeness of the cohort depends on a) eligibility criteria for inclusion, in our case the study population was geographically defined; b) initial response of the sample, the refusal rate in the Cornellà Health Interview Survey held in 1994 was about 8%; and c) the stability of the cohort on follow-up, the response rate in the follow-up was 64.3% [18]. We should consider if its necessary to enrich the cohort with more subjects (randomly selected) due to its response rate in further follow-up.

Concerning the ability to generalize results, specially in population-based cohorts, some authors have stated that in a cohort study, even if incidence rates of a disease in exposed and unexposed individuals are not externally valid, it is possible to obtain unbiased estimates of the relative risk. However, another limitation in generalizing results of a cohort study is the need for sufficient variability of exposure and outcome levels for detecting associations [18].

In conclusion, the response rate in this cohort study was relatively high. Non-participation was due to natural losses to follow-up such as migration from the geographical area of the cohort and deaths.

### Authors' contributions

EF conceived the study and with AS and MG designed the initial study protocol; EF, AS, MG, GP, CB, ES, MM and MP wrote the final study protocol; MG and AS coordinated the field work, create databases, check all data and performed statistical analysis; JMB and FJN provided expert advice in study design and data analysis; MG, EF and AS drafted the manuscript. All authors contributed to subsequent versions of the manuscript, and approved the final version. EF is the guarantor of the study.

### Competing interests

None declared.

### Acknowledgments

This study was partially funded by the Fondo de Investigacion Sanitaria (FIS/PI020261) and the Cornellà de Llobregat City Council. The authors wish to thank Francesc Saez for his technical assistance in the record linkage to Raquel Azor, Mireia Sagrera, Daniel Toro, and Elsa Plasencia for data collection, and Merce Margalef and Isaura Fite for their technical assistance. The voluntary participation of the cohort members is also acknowledged.

## References

1. Sunyer J, Lamarca R and Alonso J **Smoking after age 65 years and mortality in Barcelona.** *Spain Am J Epidemiol* 1998, **148**:575-80
2. Madans JH, Kleinman JC and Cox CS **10 years after NHANES I: report of initial follow up, 1982-84.** *Public Health Rep* 1986, **101**:465-473
3. Madans JH, Cox CS and Kleinman JC **10 years after NHANES I: Mortality experience at initial follow up, 1982-84.** *Public Health Rep* 1986, **101**:474-81
4. Davis MA, Neuhaus JM and Moritz DJ **Health behaviors and survival among middle-aged and older men and women in the NHANES I Epidemiologic Follow-up Study.** *Prev Med* 1994, **23**:369-76
5. Serdula MK, Koong S-L and Williamson DF **Alcohol intake and subsequent mortality: findings from NHANES I Follow-up Study.** *J Stud Alcohol* 1995, **56**:233-9
6. Alonso J, Orfila F, Ruigomez A, Ferrer M and Anto JM **Unmet health care needs and mortality among Spanish elderly.** *Am J Public Health* 1997, **87**:365-370
7. **The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators.** *J Clin Epidemiol* 1988, **41**:105-14
8. Riboli E and Nutrition and cancer **background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC).** *Ann Oncol* 1992, **3**:783-91
9. Grupo EPIC de España **El estudio prospectivo europeo sobre dieta cáncer y salud (EPIC) en España.** *Med Clin* 1994, **102**:781-5
10. Barton J, Bain C, Hennekens CH, Rosner B, Belanger C, Roth A and Speizer FE **Characteristics of respondents and non-respondents to a mailed questionnaire.** *Am J Public Health* 1980, **70**(8):823-5
11. Brogan Dr **Non-response in sample surveys. The problem and some solutions.** *Phys Ther* 1989, **77**:91
12. Kanuk L and Berenson C **Mail surveys and response rates: a literature review.** *J Mark Res* 1975, **12**:440-53
13. Edwards P, Roberts I, Clarke M, DiGuseppi C, Prata S, Wentz R and Kwan I **Increasing response rates to postal questionnaires: systematic review.** *BMJ* 2002, **324**:1183
14. Jacomb PA, Jorm AF, Korten AE, Christensen H and Henderson AS **Predictors of refusal to participate: a longitudinal health survey of the elderly in Australia.** *BMC Public Health* 2002, **2**:4
15. De Graaf R, Bijl RV, Smit F, Ravelli A and Vollebergh WA **Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS).** *Am J Epidemiol* 2000, **152**:1039-47
16. Stevenson J **Problems in designing and carrying out longitudinal research.** *Nord J Psychiatry* 1993, **47**:27-39
17. Hunt JR and White E **Retaining and tracking cohort study members.** *Epidemiol Rev* 1998, **20**(1):57-70
18. Szklo M **Population-based cohort studies.** *Epidemiol Rev* 1998, **20**:81-90
19. Eagan TML, Eide GE, Gulswik A and Bakke PS **Nonresponse in a community cohort study. Predictors and consequences for exposure-disease associations.** *J Clin Epidemiol* 2002, **55**:775-81
20. Lazcano E, Fernandez E, Salazar-Martínez E and Hernández-Avila M **Estudios de cohorte. Metodología, sesgos y aplicación.** *Salud Publica Mex* 2000, **42**:230-41
21. Departament de Salut Pública **Enquesta de Salut. Cornella de Llobregat, 1993-1994. Cornella de Llobregat: Ajuntament de Cornella de Llobregat.** 1995,
22. Espinàs JA, Moreno V, Borràs JM, Pujol C and Martí M **Determinants sociodemogràfics del hàbit tabàquico y de su abandono en la poblaci3n de Cornella de Llobregat.** *Gac Sanit* 1999, **13**:126-34
23. Fernández E, Schiaffino A and Martí M **Influencia del trabajo dom3stico sobre la salud y la utilizaci3n de servicios sanitarios en mujeres con trabajo remunerado y amas de casa.** *Gac Sanit* 2000, **14**:287-90
24. Martín-García AA **Conexi3n informatizada de ficheros: su uso en los registros de c3ncer de base poblacional.** *Revisiones Salud Publica* 1997, **5**:87-100
25. Fernandez E and Garcia AM **Accuracy of referencing of Spanish names in Medline.** *Lancet* 2003, **361**:351-2
26. Howe GR **Use of computerized record linkage in cohort studies.** *Epidemiol Rev* 1998, **20**(1):112-21
27. Minnick A, Roberts MJ, Young WB, Kleinpell RM and Micek W **An analysis of posthospitalization telephone survey data.** *Nurs Res* 1995, **44**:371-5
28. Galobardes B, Sunyer J, Anto JM, Castellsague J, Soriano JB and Tobias A **Effect of the method of administration, mail or telephone, on the validity and reliability of a respiratory health questionnaire. The Spanish centers of the European Astma Study.** *J Clin Epidemiol* 1998, **51**:875-81
29. Siemiatycki J and Campbell S **Nonresponse bias and early versus all responders in mail and telephone surveys.** *Am J Epidemiol* 1984, **120**:291-301
30. McHorney CA, Kosinski M and Ware JE **Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey.** *Med Care* 1994, **32**:551-67
31. Brambilla DJ and McKinlay S **A comparison of responses to mailed questionnaires and telephone interviews in a mixed mode health survey.** *Am J Epidemiol* 1987, **126**:962-71
32. Aday LA **Designing and conducting health surveys.** *San Francisco: Jossey Bass* 1989, 195-216
33. Deeg DJH **Attrition in longitudinal population studies: does it affect the generalizability of the findings?. An introduction to the series.** *J Clin Epidemiol* 2002, **55**:213-5
34. Deeg DJH, Tilburg TV, Smit JH and de Leeuw ED **Attrition in longitudinal Aging Study Amsterdam: the effect of differential inclusion in side studies.** *J Clin Epidemiol* 2002, **55**:319-28
35. Greenland S **Response and follow-up bias in cohort studies.** *Am J Epidemiol* 1977, **106**:184-7

## Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2458/3/12/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

