CENTRE FOR RESEARCH IN EPIDEMIOLOGY AND POPULATION HEALTH
EPIDEMIOLOGY OF OCCUPATIONAL AND SOCIAL DETERMINANTS OF HEALTH
INSERM UNIT 1018 & VERSAILLES-SAINT QUENTIN UNIVERSITY

THE CONSTANCES COHORT

AN OPEN EPIDEMIOLOGIC LABORATORY

SUMMARY OF THE PROTOCOL

May 2010

This document was drafted under the scientific direction of Marie Zins and Marcel Goldberg by the Cohorts team of the INSERM-UVSQ Research Unit 1018.

The complete French version of the protocol and its appendices can be downloaded from the CONSTANCES site: www.constances.fr.

16 av P. Vaillant-Couturier, F-94 800 Villejuif
Tel (33) 1 77 74 74 37 – fax (33) 1 77 74 74 03

ABSTRACT

Objectives

The main objective of the CONSTANCES project is to set up a large longitudinal epidemiologic population-based cohort intended to contribute to the development of epidemiologic research and to provide useful public health information. This cohort is intended to serve as an "open epidemiologic laboratory" widely accessible to the public health and epidemiologic research scientific community. Although designed as a "general-purpose" cohort with very broad coverage, it will focus mainly on the study of occupational and social determinants of health, on aging and women's health.

Methods

The CONSTANCES cohort is designed as a representative sample of French adults aged 18-69 years at inception covered by the National Health Insurance Fund (CNAMTS). The sample will be constructed by stratified sampling with unequal probabilities according to the variables of age, sex, and social category. The probabilities of inclusion will be empirically defined according to data from previous surveys. CONSTANCES plans to include 200 000 subjects over a 5-year period.

At inclusion, the selected subjects will be invited to fill a questionnaire and to attend a Health Screening Center (CES) for a comprehensive health examination. The individual follow-up of subjects will be both active and passive. A self-administered questionnaire will be sent annually to the subjects' homes, and they will be periodically invited to come to the CES. Work-related events and health data will be collected passively from the databases of the National Retirement Insurance Fund (CNAV), the National Death Registry (CepiDc-INSERM), and the Interfund National Health Insurance Information System (SNIIR-AM).

The data that will be collected longitudinally include social and demographic characteristics, social status, life events, behavior (e.g., smoking, alcohol, dietary habits, and physical activity), and occupational factors. The health data will cover a wide spectrum: self-reported health scales, reported prevalent and incident diseases, diagnoses of long-term chronic diseases (ALD) and hospitalizations, sick-leaves, handicaps, limitations, disabilities and injuries, healthcare utilization and services provided, death (date and medical cause). During the standardized health examination at the CES, the following data will be collected: weight, height, blood pressure, heart rate, evaluation of cognitive functions, vision, auditory, spirometry, and biological parameters (blood sugar and lipid levels, liver function tests, creatinine, complete blood counts, and urine testing). For those aged 55 years and older, a specific work-up of functional physical and cognitive capacities will be performed. A specific questionnaire will also focus on women's health. A biobank will be set up.

To take into account non-participation at inclusion and attrition throughout the longitudinal follow-up, a coefficient of correction for non-response will be determined from the analysis of the variables associated with non-participation and applied to the weight of the initial drawing attributed to each individual. To take non-participation into account, we will set up a cohort of non-participants. Passively collected data on their health characteristics, healthcare utilization, and social and occupational life course will enable to calculate a coefficient of correction for the non-response of each individual.

Ethics and data privacy

Before inclusion, different information will be provided to the eligible subjects (presentation of CONSTANCES, type of data to be collected, ability to refuse to participate, informed consent, etc.). Concrete procedures for setting up the cohorts (participants and non-participants) ensure the confidentiality of the data at every point in its circulation as well as the anonymity of the cohort of non-participants. These measures have been approved by the National Data Protection Authority (CNIL).

Schedule

The CONSTANCES project is currently in its pilot phase. Actual inclusion should start at the beginning of 2011 and take place over 5 years (2011-2015).

TABLE OF CONTENTS

GROUNDWORK FOR THE CONSTANCES PROTOCOL

Within INSERM-UVSQ Unit 1018, the Cohorts team is responsible for preparing and implementing the CONSTANCES project. The principal investigators are Marie Zins and Marcel Goldberg. The participating team members are Sébastien Bonenfant, Alain Brigand, Matthieu Carton, Mireille Coeuret-Pellicer, Sophie Launay, Mélissa Nachtigal, Anna Ozguler, Ariane Quesnot, Céline Ribet, and Rémi Sitta. Several other researchers in Unit 1018 have also collaborated in preparing the questionnaires and sampling plan.

Because of the wide range of scientific expertise necessary to prepare this protocol, we called upon numerous scientists of various backgrounds: epidemiologists specialized in the principal domains covered, biologists, health database specialists, and biostatisticians. The scientific protocol of CONSTANCES was prepared by a Scientific Steering Committee, associated with different working groups. Outside experts from France and elsewhere were also queried about specific aspects within their areas of specialization.

The questionnaire on occupational exposure was prepared in collaboration with the French Institute for Public Health Surveillance (InVS, Department of Occupational Health). The questionnaire on women's health was prepared by INSERM Unit 822. The operational protocols for the laboratory tests were prepared by the biologists of the CES of Saint Brieuc and Vandoeuvre les Nancy with the participation of the Asqualab Association. Pilot field tests were conducted in collaboration with the ClinSearch company. The data circuits and flow were organized in collaboration with CNAV, CNAMTS, and INSERM Interdisciplinary Research Institute 69 (CRI IFR 69). The procedures concerning confidentiality and personal data security were defined in collaboration with the CNIL.

The following people also contributed to this protocol:

Albert C (CNAV)
Albouy-Cossard C (CNAMTS)
Amieva H (INSERM U 897)
Andrieu S (INSERM U 558)
Ankri J (INSERM U687)
Bénézet L (DST-InVS)
Berr C (INSERM U 888),
Blondel B (INSERM U 149)
Brigand A (CES Saint Brieuc and U 1018)
Cohidon C (DST-InVS)
Colvez A (INSERM and Cetaf)
Couraud E (CES Pau)
Czernichow S (INSERM U 557 /Inra/Cnam)
Dargent P (INSERM U 149)
Dartigues JF (CHU Bordeaux and INSERM U 897)
Delcourt C (INSERM U 593)
Delmas MC (InVS-DMCT)
Dray-Spira R (INSERM U 1018)
Elbaz A (INSERM U 708)
Geoffroy-Perez B (DST-InVS)
Guéguen A (INSERM U 1018)
Hallépée S (UMS – Insee)
Hassoun D (INSERM U 822)
Henny J (CES Vandoeuvre les Nancy).
Kauffmann F (INSERM U 780)
Kuntz C (Cetaf-QEE)
L Duchet (CNAMTS)
Lang T (CHU Toulouse and INSERM U 558)
Lanoé JL (INSERM U 1018)
Leclerc A (INSERM U 1018)
Legrain S (CHU Bichat, Paris)
Lert F (INSERM U 1018)
Luce D (INSERM U 1018)
Melchior M (INSERM U 1018)
Merlière Y (CNAMTS)
Meyer JF (CES Saint Brieuc)
Moulin JJ (Cetaf)
Plu-Bureau G (CHU Necker, Paris)
Ringa V (in collaboration with N. Bajos, X. Fritel, J. Bouyer, E. de La Rochebrochard, and A. Fauconnier - INSERM U 822).
Robergeau F (CRI IFR 69, INSERM).
Roche N (CHU Hôtel-Dieu, Paris)
Roquelaure Y (CHU Angers)
Santin G (DST-InVS)
Saurel MJ (INSERM U 149)
Singh-Manoux A (INSERM U 1018)
Slama R (INSERM U 822)
Tichet J (IRSA, Tours)
Touchon J (CHU Montpellier and INSERM U 888)
Zureik M (INSERM U 700).

As well as the ClinSearch Company and the Asqualab Association

ACRONYMS

CCAM: Common Classification of Medical Acts
CepiDc-INSERM: National Death Registry
CES: Health Screening Centers
CNADS: National Consortium of Health Data Analysis
CNAMTS: National Health Insurance Fund
CNAV: National Retirement Insurance Fund
CNIL: National Data Protection Authority
CRI IFR: Interdisciplinary Research Institute 69
ICD-10: International Classification of Diseases - 10th revision
INSERM: National Institute of Health and Medical Research
InVS: French Institute for Public Health Surveillance
IreSP: Institute for Public Health Research
PMSI: National Hospital Medical Information Program
QoL: Quality of Life
RNIAM: National Registry of Health Insurance Beneficiaries
SNIIR-AM: Interfund National Health Insurance Information System
TTP: Trusted Third Party

1 OBJECTIVES OF THE CONSTANCES PROJECT

The objective of the CONSTANCES project is to set up a large population-based cohort to contribute to the development of epidemiologic research and to provide useful public health information. It is conducted in partnership with the National Health Insurance Fund (CNAMTS), the principal health insurance fund in France, which covers more than 80% of the French population, and with the Ministry of Health and the National Institute of Health and Medical Research (INSERM). This cohort is intended to serve as an "open epidemiologic laboratory" widely accessible to the public health and epidemiologic research scientific community.

CONSTANCES relies on two nationwide systems: (i) the Health Screening Centers (CES) of the CNAMTS, located across the country and equipped to collect biomedical data; (ii) CNAMTS and National Retirement Insurance Fund (CNAV) databases, which provide a permanent access to social, occupational and health data.

1.1 A TOOL FOR EPIDEMIOLOGIC RESEARCH

CONSTANCES is designed to be a large cohort with a number of subjects, quality and diversity of data, and follow-up methods comparable to the largest international cohorts. This should allow to build a powerful tool for epidemiologic research in France.

The scientific objectives of CONSTANCES are focused largely on the epidemiology of occupational and social health determinants, but the cohort should also make it possible to conduct projects on a variety of epidemiologic themes, due to the open access provided to the community of public health and epidemiology researchers: CONSTANCES is designed as an open epidemiologic laboratory and will make a major contribution to public health research in France.

1.2 A TOOL FOR PUBLIC HEALTH

Numerous data sources in France can provide public health officials with the information they need. Nonetheless these all have limitations related to the field they cover and to the nature and quality of the data. Among these limitations, one often highlighted is the lack of large longitudinal studies covering a broad health field and its determinants, taking into account the interaction of numerous factors as well as the changes over time of people and the socioeconomic environment (Valleron, 2006).

CONSTANCES was designed mainly as a tool to support the public health objectives of CNAMTS and of the national government, by its thorough system for the follow-up and collection of very diverse information through a variety of methods and data sources about a large representative sample of the adult population.

1.3 GENERAL ORIENTATION: AN "OPEN EPIDEMIOLOGIC LABORATORY"

The CONSTANCES cohort will be a large sample, representative of the general population, characterized by broad coverage of health problems and health determinants and openness towards diversified users. It is designed to be an open epidemiologic laboratory. It will serve as an important scientific instrument, rather like a telescope or a particle accelerator, for example, or a genotyping laboratory equipped with a sequencer — built not to answer a specific question but rather to help analyze a wide range of scientific problems and will be accessible to the community of specialized researchers.

For this reason, CONSTANCES has no specific objectives in terms of hypotheses about specific diseases and/or risk factors. For the same reason, the duration of the project has not been defined: the cohort that will be constructed is intended to be the object of longitudinal follow-up without any time limitations, able to study the effects of risk factors in the very long term. CONSTANCES will thus be able also to take into account the progress in knowledge and techniques that will continuously raise new scientific questions about which CONSTANCES will be able to provide enlightenment.

1.4 SCIENTIFIC THEMES

Although designed as a general-purpose cohort with a very broad scope, its major orientation is the study of occupational and social determinants of health. These themes are essential fields of research in public health and epidemiology today. They involve numerous health problems and diverse populations.

Occupational factors

Epidemiologic knowledge of the various effects of occupational factors is needed: effects of occupational exposure and working conditions, role of occupational and social life course, analysis of occupational determinants of health inequalities, etc.

There are numerous health risk factors of occupational origin: chemical hazards, noise, temperature, vibrations, radiation, biological agents, physical and postural constraints, mental load and stress, hours, and work pace. They concern a very large fraction of the population. Psychosocial factors linked to the organization of work and to an imbalance between the individual's efforts and rewards are also sources of potentially pathogenic stress (Karasek & Theorell 1990, Siegrist 2002). Numerous studies have shown the role of psychosocial factors at work in cardiovascular disease, in the incidence of mental disorders, including depression, on quality of life (QoL), and in musculoskeletal diseases (Marmot et al., 1998). The way that companies are run, including, for example, the increasingly frequent use of subcontractors, fixed-term contracts, and temporary workers, creates different processes that increase precarity and affect living conditions. Little is known about the relations between these processes and health.

Working conditions, occupational exposures, and life-long occupational trajectory are major determinants of the aging process. The disorders that lead over time to impairments, disabilities, and diseases most often originate early in working life and build up over time until they become chronic. The same is true of occupational constraints and hazards. All may lead to an early exit from the labor market.

Finally, occupational factors are also major determinants of social inequalities in health.

Social determinants of health and social inequalities

In France, the system of universal access to health care corresponds to an ideal of equality in the face of disease and death. Nonetheless, while health status improves generally, social inequalities in health continue, and some even worsen. Moreover, as research has shown, this phenomenon is not limited to the degradation of health status among the most disadvantaged groups. Rather, there is increasing evidence of an inverse gradient between various measures of socioeconomic position and diverse health problems (Marmot et al., 1991). These observations have renewed questions about the causes of these inequalities (Leclerc et al., 2000) and about their public health implications; if policies to combat these health inequalities are limited to categories described as "insecure" or disadvantaged, they will be globally ineffective.

From a public health point of view, developing and implementing policies aimed at reducing social inequalities in health requires a better description of the distribution of health problems of populations characterized by diverse criteria of social status and a better understanding of the nature of the determinants of the disparities observed and the pathways by which they work (Marmot et al. 2008). Epidemiologic analysis of the social and occupational determinants of health is therefore a major issue, from both scientific and public health perspectives.

Aging

Epidemiologic data remain sparse on the topic of changes in health with age and more particularly about aging and its relation to health, work, and life course. Studies are essentially limited to the age groups above 60 years and provide little information about earlier life periods (Dartigues & Alpérovitch, 2005), even though factors that lead to impairments, disabilities, and chronic diseases at advanced ages often begin early in life, and they continue to accumulate throughout life. We still know little about social inequalities in the aging process or about the respective roles played by individual susceptibility

factors — especially genetic factors, life course, and living conditions in all their dimensions, the conditions in which people stop working, their attention to their health, and primary, secondary and tertiary prevention.

The ageing-related scientific objectives in CONSTANCES involve the study of the role of risk factors of various types throughout life. Another objective is to document the determinants of health most frequently encountered or suspected in chronic age-related diseases. Longitudinal follow-up offers broad possibilities for dynamic study of the delayed effects of working conditions and occupational exposure on aging (e.g., frailty, cancer, chronic diseases, and mental health), the factors that may lead to inactivity and isolation, factors and mechanisms that contribute to successful aging, or on the contrary are at the origin of disabilities and/or frailty.

Women's health

Women's health presents its own specific issues, in particular in terms of reproductive history, the hormonal system, sexuality, and contraception. Moreover, some diseases are especially frequent among women (osteoporosis, urinary incontinence, etc.). Specific data on women's health will be collected, to examine in depth the specific aspects. Until now, the social inequalities related to women have been inadequately understood and badly measured. CONSTANCES should allow us to study not only fields specifically related to women in terms of health, but also to have a more complete vision of women's social situation (gender relationships, social role, etc.) as it is related to health.

2 METHODS: ESSENTIAL ELEMENTS OF THE PROTOCOL

2.1 ESTABLISHMENT AND FOLLOW-UP OF THE COHORT: OVERVIEW

CONSTANCES is a prospective epidemiologic cohort, with an undefined duration of follow-up. The sample that constitutes the cohort will be representative of the population covered by the principal national health insurance fund (CNAMTS) and aged 18 to 69 years at inclusion; the target number is 200 000 subjects, and its structure will be proportional to the general population for sex, age, and social category.

The principal stages of the establishment and follow-up of the cohort, which will be detailed in the sections below, are summarized briefly here.

Selection of eligible subjects: eligible subjects will be randomly drawn by stratified sampling with unequal probabilities. We will seek to overrepresent individuals least likely to volunteer, according to the standard variables (age, sex, and social category). The CNAV will conduct the random drawings from their databases, which include the entire French population.

Invitation to participate: This will include an invitation to come for a health examination at a CES. The randomly selected subjects will receive a letter describing the CONSTANCES project and a brochure presenting CONSTANCES and the CES, as well as a reply form that they can return to indicate their consent to participate in the cohort as part of the health examination.

Inclusion of subjects volunteering for the cohort: Those who agreed to participate in CONSTANCES will receive an appointment to come to their CES at a specific time, day, and place, as well as a self-administered questionnaire to be completed at home.

Data collection at inclusion: Besides the self-administered questionnaire, completed at home, the subjects will have a comprehensive health examination to collect the following health data: clinical examination, complete blood count, blood pressure measurement, weight, waist and waist/hip ratio, electrocardiogram and spirometry, eye and ear examinations, additional questionnaires (one administered face-to-face about lifetime occupational exposure, one self-administered to complete onsite, for the women). Those who agree to participate at the end of this first examination will be asked to sign an informed consent form if they agree to participate.

Active follow-up: A self-administered questionnaire will be sent to the subjects at their homes each year; an examination at the CES every 5 years is planned for all cohort members.

Passive follow-up of work-related events and health data: The principal work-related events will be regularly extracted from the databases of the CNAV, which records them continuously. The health data will include vital status and causes of death (CepiDc-INSERM database), as well as the principal health events, which will be extracted from the Interfund National Health Insurance Information System, SNIIR-AM (data about health care reimbursement and long-term diseases), from the national hospitalization database (principal diagnosis and associated diagnoses, diagnostic procedures, and treatment for each hospitalization).

2.2 COHORT COMPOSITION

Source population and cohort structure

The source population is that of the people in France whose health insurance is administered by the CNAMTS. This fund covers more than 80% of the French population, that is, approximately 50 million people.

At inclusion, the CONSTANCES cohort will be a representative sample of the general French adult population aged 18 to 69 years and insured by the CNAMTS. It will be representative for the variables of age, sex, and social category.

Resource facilities: Health Screening Centers

Everyone with health insurance from CNAMTS, as well as their dependents, is entitled to receive health examinations that include free very complete work-ups. Overall the CES conduct approximately

600 000 health examinations annually. CONSTANCES subjects will be included in 17 CES distributed throughout France that have experience with the recruitment of large numbers of people and have the necessary quality of high-level technical equipment.

Number of subjects – Power

To be able to answer the many questions raised in varied domains, CONSTANCES must be a large sample. It is impossible to carry out the standard calculations of power and necessary sample size that are performed for studies with more specific objectives. It is nonetheless important to assess the potential of CONSTANCES in terms of its capacity to conduct epidemiologic studies likely to have good statistical power. Accordingly, we estimated the number of health events expected in the CONSTANCES cohort over a more or less long term in a cohort with an age and sex structure identical to that of the French general population aged 18 to 69 years at the 1999 census. We calculated the number of expected events at the end of 5, 10, and 15 years for events for which we have reliable national reference data: deaths and incidence of cancer, ischemic heart disease and Alzheimer disease. We note that the number of these serious events is high and will make possible numerous studies with good power.

Expected number of health events

(Reference rates: CepiDc 1999; Remontet *et al*.2002; ARME 2007)

	5-year follow-up			10-year follow-up			15-year follow-up		
	M	W	Total	M	W	Total	M	W	Total
Death, all causes	4131	2133	6264	9727	5502	15 229	16 983	10 736	27 719
Incident cancers	3162	2220	5381	7036	4855	11 892	11 444	7823	19 267
Ischemic heart disease (35-64 years)	681	138	819	1418	290	1708	2178	452	2630
Alzheimer disease	265	240	505	793	1007	1800	1548	2469	4018

The subjects lost to follow-up thus present another problem affecting the power of a longitudinal study (Goldberg & Luce 2001). It is impossible to estimate precisely the number of subjects who will be lost to follow-up in the CONSTANCES cohort over the years. We can nonetheless makes estimates based on experience with the follow-up of other French cohorts, such as Gazel, which began in 1989 with more than 20 000 subjects (Goldberg et al. 2007). Active participation by self-administered questionnaire is high: at the end of 19 years, only 3.1% of the subjects who participated at inclusion never returned an annual questionnaire. The number of subjects really lost to follow-up, that is, whom we can no longer locate in the databases, is tiny: approximately 0.9% after 19 years of follow-up. It is reasonable to think that CONSTANCES, which will apply similar methods, will also have an efficient follow-up.

Duration of follow-up

Because the advantages of longitudinal follow-up increase with its duration and in view of the broad objectives for this cohort, the duration of follow-up should be as long as possible. Its planned duration is therefore indefinite.

To the extent that the initial population will age as the follow-up continues, it will be necessary to renew the participants to maintain over time a social and demographic structure comparable to that of the source population. It is nonetheless too early to consider the procedures for renewing the cohort. This topic will be discussed in several years.

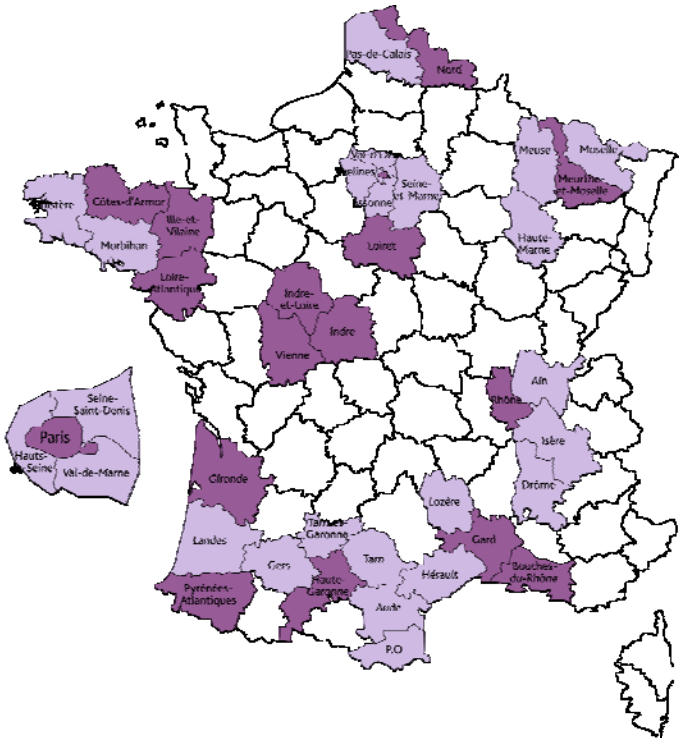
2.3 PROCEDURES FOR INCLUSION

Participating CES

Currently 17 CES have agreed to participate in the project. All are large, have a staff motivated to work in epidemiology, and use advanced equipment; their geographic distribution represents the principal

regions of France. The participating CES have sites in 16 different districts but their actual catchment area is wider, as Figure 1 shows.

Figure 1: The participating CES in France



The dark colors represent districts where CONSTANCES CES are located,
the lighter colors, the other districts they cover

Duration of inclusion

We will proceed gradually to the inclusion of the entire cohort over a 5-year period. Each wave will include 40 000 subjects in a year. The number is realistic in view of the volume of subjects seen and examinations performed by the participating CES. The final cohort will be constituted at the end of this 5-year period.

Procedures

Randomly selected eligible potential subjects will be asked to participate in CONSTANCES. The random drawing methods and the practical procedures used for inclusion are described below.

2.4 PROCEDURES FOR THE LONGITUDINAL FOLLOW-UP

Tracing subjects

Post office procedures make it possible to obtain regular updates of participants' postal addresses. Should this method fail, other databases will be used. If they provide no useful results, the CES of inclusion will search for the individual's address.

Subjects' active participation in follow-up

An annual self-administered questionnaire will be sent to the subjects at home. Maximizing their personal participation rate is essential. Accordingly, regular contact with participants will include a CONSTANCES Cohort Journal, which will present results, associated projects, etc., and will be sent regularly to participants. A website will also be created.

Passive collection of work-related events and health data

The subjects included in CONSTANCES will also be followed up "passively" (so-called because this follow-up will not require the subjects' participation) for work-related events and health data by regular linkage with the national databases.

Work-related events

The CNAV databases are essential for access to work-related data. This agency's role is to ensure the rights to pension payments after retirement for every individual in France who had health insurance from CNAMTS at least once during his or her life. It has therefore set up a system that allows it to collect social data from different organisms and schemes that manage various forms of insurance and other social protection. The CNAV regularly receives for its databases employers' annual reports, and information about periods of employment and unemployment from organisms of social protection (e.g., sick leave, maternity leave, and diverse social benefits).

Health data

Vital status and causes of death will be obtained from CepiDc-INSERM, which manages the national database of causes of death.

Access to the Interfund National Health Insurance Information System (SNIIR-AM), which covers the entire French population, should be an efficient method of obtaining information about health events. The advantage of SNIIR-AM from an epidemiologic perspective is that it contains individual medical data, structured, and coded in a standardized manner.

It is made up of several databases. The health insurance reimbursement database is appropriate for analyzing prescribing practices but does not include direct information about the nature of the diseases treated. By definition it excludes self-medication and services for which no reimbursement is sought. The long-term disease (ALD) database covers all insurance beneficiaries exempt from co-payments and user fees: it codes the exempting disease according to the International Classification of Diseases 10th revision (ICD 10). The Hospital Medical Information Program (PMSI) national database includes the following information for each hospitalization: the principal and any associated diagnoses, age, sex, and the most expensive diagnostic and treatment procedures. The diagnoses are coded according to ICD 10 and procedures according to the Common Classification of Medical Acts (CCAM).

2.5 PRINCIPAL DATA TO BE COLLECTED FROM DIFFERENT SOURCES

Domains covered

The data to be systematically collected for all participants are intended to build a corpus that allows to describe and follow trends over time. They will constitute both a database and a reference source for health status, general morbidity and mortality, socioeconomic and occupational status, the familial, social, and geographic environment, and personal and environmental risk factors.

Data

Generally, we selected variables already used in other surveys, both because they are validated measures and because it will thus be possible to have reference data for some analyses. Whenever it was possible and pertinent, we used scales already published in the literature, for which the psychometric properties are already established. Below, we present the type of data that will be collected.

Social and demographic characteristics, social status, and situation

Work status and situation, educational level, income level, marital status, household composition, socioeconomic status of parents and spouse, and material living conditions (type of housing, household income, etc).

Health data

Personal and family history (cancer, cardiovascular, psychiatric), self-reported health scales (perceived health, QoL, mental health, and specific scales for cardiovascular, musculoskeletal, and respiratory diseases), diseases (list of reported diseases, ALD and hospital diagnoses, sick leaves, handicaps, limitations, disabilities and injuries and healthcare utilization and management), and date and cause of death.

In the CES: weight, height, waist-hip ratio, blood pressure, heart rate, vision, hearing, and lung function, laboratory tests (blood sugar level, lipid work-up, liver function tests, blood creatinine levels, complete blood counts, urine tests).

A Biobank will be set up; its preparation is currently underway.

Behavior

Smoking and alcohol consumption (past and present), dietary habits and physical activity, marijuana use, and sexual orientation.

Occupational factors

Job history, lifelong and current occupational exposure to chemical, physical, and biological agents, postural, mechanical and organizational constraints, and stress at work.

Specific health problems of the elderly (55 years and older)

Evaluation of functional capacities: IADL (Instrumental Activities of Daily Living) scale (Lawton et al, 1969), questions from the French Handicaps-Disabilities-Impairments Survey (DREES, 2001), ability to use new technologies.

CASP (Control, Autonomy, Self-realisation and Pleasure; Hyde et al., 2003) QoL scale, particularly appropriate for senior citizens.

Cognitive functions: MMSE (Folstein et al., 1985); Trail Making Test A - B (Boll & Reitan, 1973 Miner et al, 1998); Wechsler's coding subtest (Wechsler, 1981); Digital Finger Tapping Test (Mitrushina et al. 1999); Word fluency, formal lexical and semantic evocation (Borkowski et al., 1967, Cardebat et al. 1990); Grober & Busckhe's memory test (Grober et al. 1998; Van der Linden et al. 2004).

Physical functioning: Gait Speed Test (Shkuratova et al., 2004); Balance Test: standing on one foot for 10 seconds (Horak et al., 1989); Hand Grip Test (Giampaoli et al., 1999).

Health problems specific to women

Treatment for menopause, osteoporosis and osteoporotic fractures, sphincter and static perineal disorders, benign breast disease, endometriosis and chronic pelvic pain, infertility and delayed childbearing, sexually transmitted diseases, and issues related to sex life.

Periodicity of data collection

The periodicity of follow-up will vary according to the sources.

The self-administered mail questionnaire will be sent annually, thus allowing close follow-up, by collecting numerous data without asking subjects for too much work. At the same time, it will facilitate rapid reaction for setting up new studies and the establishment of a sense of loyalty in the participants; too long a delay between two questionnaires is a factor that promotes dropping out. Some data will be collected each year (health status and reported morbidity, life events and characteristics of place of residence, smoking, alcohol, etc.); others will be collected at longer intervals, according to a planned calendar (health scales and questionnaires for a specific health area or specific risk factors.) The mailing of self-administered questionnaires will be staggered over the year to take seasonal variations into account, since they are important for some topics (in particular, morbidity, drug use, and working conditions).

Because the national databases essentially record events continuously, the follow-up of the data they provide will be permanent.

Participants will also be asked to come to the CES every 5 years for medical and laboratory examinations. The periodicity may vary according to specific groups (more frequently for the elderly) or scientific questions.

2.6 QUALITY CONTROL AND VALIDATION OF HEALTH EVENTS

The self-administered questionnaires will undergo the standard verifications: percentages of non-response, missing data, delay in return, etc.

For the data collected during the inclusion visit to the CES, routine permanent quality control is planned to assess the accuracy, reproducibility, concordance, and internal and external validity of the data collected and to study their factors of variability. Epidemiologic research assistants will make regular onsite inspections.

For the data extracted from the national databases, particular attention will be paid to validation of the diagnoses extracted from the health-related administrative databases, which will be routinely verified. Initially, two types of disease will be analyzed most particularly: ischemic cardiovascular events and cancers reported in the available sources will be routinely verified at the hospitals or with the general practitioners. Moreover, collaboration with different research teams is being set up as part of the National Consortium of Health Data Analysis (CNADS) to develop methods for analysis of routinely collected health-related administrative data for epidemiologic and health evaluation applications; this project is funded by the Institute for Public Health Research (IReSP).

2.7 REPRESENTATIVENESS AND SELECTION EFFECTS

One of the major sources of bias in epidemiologic surveys comes from selection effects; it can bias estimates of disease prevalence or incidence (or of prevalence of exposure to a risk factor) and of associations between exposure and disease. In longitudinal cohorts, selection effects may occur at inclusion and throughout follow-up because of cohort attrition (Goldberg & Luce, 2001). The problem of bias linked to selection effects is very different depending on whether the objectives are analytic or descriptive.

In a cohort whose inclusion procedures are the same for all subjects (the case of CONSTANCES), in principle the exposure-disease relation does not differ between subjects who are included and those who are not (Criqui, 1979; Austin et al., 1981). In principle, therefore, the selection procedures for CONSTANCES participants should generate minimal bias, if at all, in analytic studies. On the other hand, the problem of attrition during follow-up may cause quite substantial bias if the probability of continued follow-up is different in exposed and unexposed subjects or in those who do or do not become ill, and this is often the case.

For descriptive studies of the frequency of health problems and exposures, the indicators of interest must be estimated in a representative sample of the target population. We have verified that the structure of the population of the districts where the CONSTANCES CES are located is essentially identical to that for France as a whole for the principal demographic, social, and occupational characteristics; we should thus be able to generalize the CONSTANCES results to the French population as a whole.

Using volunteer subjects inevitably produces selection effects, even in studies that use random drawing from an appropriate sampling base. At inclusion, there are non-participants, a potential source of bias. To compensate, researchers usually attempt to collect a minimum data set for the non-participants (mainly age, sex, and social category), to facilitate subsequent adjustments for estimating the relevant parameters. This approach nonetheless has some limitations. First, it is not always possible to collect the adjustment data for non-participating subjects. Nor is it always clear whether these data are sufficient to control for potential bias, because we know, for example, that within the same socioeconomic category there are many important differences in terms of health, behavior, lifestyles, social networks, etc. (Goldberg et al. 2001). Finally, it is rarely possible to control completely for potential selection bias because it is rare to have the relevant data collected simultaneously for the participants and the non-participants. To obtain a representative sample of the target population and to minimize the bias associated with selection effects at inclusion and during follow-up in CONSTANCES, we plan to take the steps described below.

Sampling base

The sampling base at inclusion is composed of all persons aged 18 to 69 years and covered by CNAMTS in the catchment areas of the 17 CONSTANCES CES.

Random drawing

The random drawing will be stratified according to unequal inclusion probabilities, based on data from participation in surveys involving invitations to the CES: the 2002-2003 French Decennial Health Survey and the 2004 Prevalence survey for hepatitis B and C (InVS 2005).

Adjustment for non-participation

We will set up a sample of non-participants for whom we have data on social and demographic characteristics, through the CNAV files (sex, age, work status, social category), as well as information about health and health-care utilization from SNIIR-AM. As we will have data from the SNIIR-AM and CNAV files for both the participants and the sample of non-participants, we will be able to estimate the probabilities of participation in CONSTANCES with prediction models (logistic regression or segmentation tree); the inverse of the probability of participation will then provide the adjustment coefficient for each participant.

Adjustment for attrition

We can assume that almost none of the people included in CONSTANCES will be permanently lost to follow-up, since the participants will be followed passively through the SNIIR-AM and CNAV files. There will nonetheless be attrition due to the failure to return the annual questionnaire. Thus, adjustment for attrition is necessary if the analyses of the variables in the questionnaire are to be valid.

Inclusion in CONSTANCES will take place over a 5-year period. For wave 1, we will distinguish the participants who returned the self-administered questionnaire in year 2, that is, the year following their inclusion (which is year 1), from those who did not. We will have the data collected at inclusion in CONSTANCES (year 1) for all participants as well as the SNIIR-AM and CNAV data corresponding to year 1 or to year 0 (the year before inclusion). The coefficients of adjustment for attrition in year 2 can thus be calculated by a method similar to the one used to calculate the coefficient of adjustment for non-participation. For the following years, we will again distinguish the participants who returned the self-administered questionnaire from those who did not.

It is likely, however, that some number of CONSTANCES participants will not respond one year but will again respond in the following years. Methods to deal with this can be derived from partial non-response methods: exclusion of non-respondents, use of indicator variables for missing data, simple or multiple imputations, or re-weighting. The choice of the most appropriate method must be defined on a case-by-case basis according to the specific problem being studied.

Passive and longitudinal follow-up of participants in the SNIIR-AM and CNAV databases, whether or not they drop out of the cohort by not returning the annual questionnaire, will make it possible to update the coefficients of adjustment for attrition.

Sample balancing

Each year, we will adjust the cohort data on the reference population. The first year, we will randomly draw from the CNAV files a sample of CNAMTS members in the CONSTANCES districts aged 18 to 69 years. The second year, and respectively the third, fourth and fifth years, this sample of randomly drawn individuals will include people aged 18 to 70 years, then 18 to 71 years, 18 to 72 years and 18 to 73 years. Each of these samples will be twice as large as that of the population of CONSTANCES participants so that the reference population will be significantly greater than the sample.

After fusion of this file with the SNIIR-AM files, we will calculate the relevant margins and thus, beyond socioeconomic and demographic characteristics, be able to integrate the variables relative to the health and healthcare utilization characteristics (information also available for the survey participants). The quality of weighting will be therefore substantially improved by the calculation of margins specifically related to health, the focus of the cohort.

2.8 OPERATIONAL ASPECTS OF INCLUSION AND FOLLOW-UP

Preliminary information campaigns before inclusion in CONSTANCES

Before the random drawing, a general public information campaign will take place in the CONSTANCES areas, via local media and the CNAMTS sites. Subjects' right to refuse to participate in the study will be mentioned.

Constitution of cohorts (participants and non-participants)

1) The CONSTANCES team will provide to CNAV the software to conduct the random drawing; the probabilities of participation are empirically defined on the basis of surveys already conducted among random samples asked to come for a CES examination.

2) The CNAV will extract, via the comprehensive National Registry of Health Insurance Beneficiaries (RNIAM), the persons insured by the funds covering the CONSTANCES CES, link them to social category data and perform the random drawing. Each randomly drawn subject will receive an individual non-identifying CONSTANCES number. The CNAV will address to a "Trusted Third Party" (TTP) and to the CNAMTS the necessary data for the following operations.

3) The CNAMTS will generate a new non-identifying individual number, different from the preceding one, for linking to its databases.

Invitations

The TTP will send to a printer-shipper the data for addressing the invitations and sending them to the randomly-selected subjects. A response-coupon will be enclosed; they can use it to indicate their agreement to participate in the cohort or their refusal to participate either in the principal study or as one of the non-participants followed passively in the national databases.

All documents will specify the cohort's aims and procedures. The documents will provide information about the treatment of their data, and subjects have a right of access to their data, including the right to have the personal data concerning them modified, corrected, and deleted. CONSTANCES has, in compliance with French laws, the legal authorization from CNIL, which is responsible for ensuring the confidentiality of personal data.

Appointments and questionnaires to be completed at home

The people who agree to participate in the cohort will receive a letter with an appointment at CES, together with an explanatory letter and a questionnaire to complete at home and bring with them to the CES.

Inclusion of participants by the CES and data collection

Questionnaires

In addition to the questionnaire completed at home, three other questionnaires will be completed during the CES visit: a self-administered questionnaire on women's health, a questionnaire on occupational exposures (administered by CES staff), and a questionnaire completed by the physician at the end of the clinical examination.

Medical examination

The procedures are set by standard operating procedures to guarantee strict standardization. The following examinations shall be performed: biometry, vision, hearing, spirometry (lung function), ECG, blood pressure, and functional examinations for those 55 years and older. Quality control procedures shall be regularly implemented.

At the end of the CES visit, the participants willing to continue in the cohort will sign an informed consent.

Data transmission

The data from the health examination results and from the questionnaires completed at home and at the CES, identified by the individual non-identifying CONSTANCES number, will be sent by each CES to the CONSTANCES team.

Data storage

Once computerized, the data will be associated with the cohort member's permanent non-identifying number (NCP) and included in the project databases on a server accessible only to authorized personnel.

Passive follow-up (linkage to national databases)

CNAV data

For each inclusion wave, the CONSTANCES team will send to the CNAV a list of the individual non-identifying CONSTANCES numbers and the corresponding participation status. The CNAV will extract from its information system the vital status and work-related data, which it will send to the CONSTANCES team. As part of the follow-up, the CNAV will periodically forward to the CONSTANCES team new work-related and vital status data for participants and non-participants.

CNAMTS data

For each inclusion wave, the CONSTANCES team will send to the CNAMTS the list of individual non-identifying CONSTANCES numbers and the corresponding participation status. The CNAMTS will apply an algorithm to generate an encrypted number permitting access to SNIIR-AM for each participant and will extract from its information system the medical data, which it will send to the CONSTANCES team. As part of the follow-up, the CNAMTS will periodically forward to the CONSTANCES team new work-related and vital status data for participants and non-participants.

Deletion of nominative data

After inclusion, the Trusted Third Party will destroy all data that do not concern the participants (including for those belonging to the non-participant cohort, whose follow-up will be strictly anonymous). Accordingly, after inclusion, the Trusted Third Party will have available only data about the participants, to ensure their follow-up.

Causes of death

The causes of deaths among the cohorts of participants and non-participants will be collected according to the French legal procedures which authorizes access to death data for health research. The access involves two successive stages: (1) research of vital status, to learn if the person has died; (2) research of the causes of death in the CepiDc-INSERM database. The CNAV plays a role in this circuit by providing nominative data about participants and non-participants to the Trusted Third Party, who will conduct the necessary searches.

Active follow-up — Querying the participants by self-administered questionnaires

On the anniversary date of inclusion in the cohort, each participant will receive a self-administered follow-up questionnaire, accompanied by a follow-up form intended to update the nominative data (change the name they use, or their address).

Mailing of questionnaires

The CONSTANCES team will send to the Trusted Third Party the list of individual non-identifying CONSTANCES numbers to be contacted. The Trusted Third Party will use the nominative database to send to a printer-shipper the data necessary for the mailings.

Treatment of the returned self-administered questionnaires and follow-up forms

The CONSTANCES team will receive all of the self-administered follow-up questionnaires and treat them: scanning, recognition, videocoding, filing, etc. At the end of this treatment, the data will be included in the CONSTANCES database.

Management and treatment of addresses

After participants are included at the CES, the Trusted Third Party will hold the data that make it possible to write to them. The CONSTANCES team will generate a specific "Post Office" transfer number to ensure correspondence between this number and the individual non-identifying CONSTANCES number, which it will transmit to the Trusted Third Party. The TTP will send a file to the Post Office so that the latter can normalize the addresses, correct addresses for those who have moved, and enrich the file (including geocoding = X, Y, census block code).

During follow-up, the annual questionnaires will be mailed according to a system established by the Post Office, which will facilitate the rapid identification of wrong addresses.

3 SCHEDULE

The pilot phase of the CONSTANCES project is currently completed. It included two main parts: partial pilot tests of the procedures at the CES (questionnaires, medical and other examinations, etc.), and a full-scale test of about 3,500 subjects, from a random drawing of subjects asked to participate up to the data collection stage at seven CES (Lille, Bordeaux, Saint Brieuc, Tours, Pau, Toulouse, and Rennes), and their linkage with the national databases.

Actual inclusion should start at the beginning of 2011 and take place over 5 years (2011-2015).

REFERENCES

ARME. Incidence et prévalence de différentes maladies – Bordeaux, ARME Pharmacovigilance – 2007.

Austin MA, Criqui MH, Barrett et al. The effect of response bias on the odds-ratio. Am J Epidemiol, 1981;114,137-143.

Boll TJ, Reitan RM. Effect of age on performance of the Trail Making Test. Percept Mot Skills 1973;36:691-4.

Borkowski JG, Benton AL, Spreen O Word fluency and brain damage. Neuropsychologica 1967; 5:135-140.

Cardebat D, Doyon B, Puel M, Goulet P, Joanette Y. Formal and semantic lexical evocation in normal subjects. Performance and dynamics of production as a function of sex, age and educational level. Acta Neurol Belg. 1990; 90: 207 – 17.

Criqui MH. Response bias and risk ratios in epidemiologic studies. Am J Epidemiol, 1979;109,394-399.

Dartigues JF, Alpérovitch A. Épidémiologie et vieillissement. In : Valleron AJ (ED.) : « épidémiologie : conditions de son développement, et rôle des mathématiques ». Rapport scientifique et technique. Paris, Académie des Sciences, 2005 (sous presse).

Folstein M, Anthony JC et al. The meaning of cognitive impairment in the elderly. J Am Geriatr Soc 1985;33:228-35.

Giampaoli S, Ferrucci L, Cecchi F, et al. Hand-grip strength predicts incident disability in non-disabled older men. Age and Ageing 1999; 28: 283-288.

Goldberg M, Chastang JF, Leclerc A, Zins M, Bonenfant S, Bugel I, Kaniewski N, Schmaus A, Niedhammer I, Piciotti M, Chevalier A, Godard C, Imbernon E. Socioeconomic, demographic, occupational and health factors associated with participation in a long-term epidemiologic survey. A prospective study of the French Gazel cohort and its target population. Am J Epidemiol 2001;154:373-84.

Goldberg M, Leclerc A, Bonenfant S, Chastang JF, Schmaus A, Kaniewski N, Zins M. Cohort profile: the GAZEL Cohort Study. Int J Epid. 2007 36:32-39.

Goldberg M, Luce D. Les effets de sélection dans les cohortes épidémiologiques. Nature, causes et conséquences. Rev Epidemiol Santé Publique 2001;49:477-92.

Grober E, Buschke H, Cystal H et al. Screening for dementia by memory testing. Neurol., 1998; 38:900-903.

Horak FB, Shupert CL, Mirka A. Components of postural dyscontrol in the elderly: a review. Neurobiol Aging. 1989;10: 727-38.

Karasek R, Theorell T. Healthy Work: stress, productivity and the reconstruction of working life. New York: Basic Books, 1990.

Marmot MG, Davey Smith G, Stansfeld S et al. Health inequalities among British civil servants: the Whitehall II study. Lancet 1991;337:1387-92.

Marmot MG, Fuhrer R, Ettner SL, Marks NF, Bumpass LL, Ryff CD. Contribution of psychosocial factors to socio-economic differences in health. The Milbank Quarterly 1998; 76.

Marmot M, Friel S, Bell R, Houweling T, Taylor S. Closing the gap in a generation: health equity through action on the social determinants of health. Lancet 2008; 372: 1661–69.

Mitrushina MN, Boone KB, D'Elia LF. Handbook of Normative Data for Neuropsychological Assessment. New York: Oxford University Press. 1999.

Siegrist J. Effort-reward Imbalance at Work and Health. In: Research in Occupational Stress and Well Being, Historical and Current Perspectives on Stress and Health. In: P. Perrewe (Eds). JAI Elsevier, London. Vol. 2, pp 261-291, 2002.

Remontet L, Buemi A, Velten M, Jougla E, Estève J. Évolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000. Saint Maurice, InVS, 2002.

Shkuratova N, Morris ME, Huxham F. Effects of age on balance control during walking. Arch Phys Med Rehabil. 2004;85(4):582-8.

Valleron AJ (Ed.) : « épidémiologie : conditions de son développement, et rôle des mathématiques ». Rapport sur la Science et la Technologie n° 23, Comité RST de l'Académie des sciences. Éditions EDP Sciences, 2006.

Van der Linden M, Coyette F, Poitrenaud J, Kalafat M, Calicis F, Wyns C, Adam S et les membres du GREMEM. L'épreuve de rappel libre/rappel indicé à 16 items (RL/RI-16. L'évaluation des troubles de la mémoire. Van der Linden M et les membres du GREMEM (eds). Solal éditeur, Marseille - 2004.

Wechsler D. Manual for the Wechsler Adulte Intelligence Scale-Revised. New-York: Psychological Corporation, 1981.