

The **CONSTANCES** cohort

<https://www.constances.fr/>

Summary of the protocol (Updated May 2022)

UMS 011 - Cohortes épidémiologiques en population

16 av P. Vaillant-Couturier, F-94 800 Villejuif

Tel (33) 1 77 74 74 37 – fax (33) 1 77 74 74 03

TABLE OF CONTENTS

1	OBJECTIVES OF <i>CONSTANCES</i> : A GENERAL POPULATION-BASED COHORT	3
2	METHODS: MAIN ELEMENTS OF THE PROTOCOL	3
2.1	Setting up and follow-up of the cohort: overview	3
2.2	Who is in the cohort	4
2.3	Main data collected	5
2.4	Data sources	6
2.5	Volunteer recruitment and data collection at inclusion	7
2.6	Follow-up	8
2.7	Selection effects and representativeness	9
2.8	Regulatory issues	10
2.9	References	10
2.10	Publications from <i>Constances</i>	11
3	ACCESS TO THE <i>CONSTANCES</i> COHORT	11

1 OBJECTIVES OF *CONSTANCES*: A GENERAL POPULATION-BASED COHORT

The objective of the *CONSTANCES* project is to implement a large population-based epidemiological cohort designed to contribute to the development of health research and to provide information for public health purposes. Carried out in close partnership with the National Health Insurance Fund (CNAM), which provides most of the funding, and with the National Old Age Insurance Fund (CNAV), the project concerns the adult population affiliated with the Social Security, which covers more than 85% of the French population. *CONSTANCES* is a "National Infrastructure for Biology and Health", widely accessible to the public health and research community. The "Population-Based Cohorts" unit (UMS 11), a joint project between the University of Paris Cité, INSERM, the University of Paris Saclay and the University of Versailles St Quentin is in charge of this project.

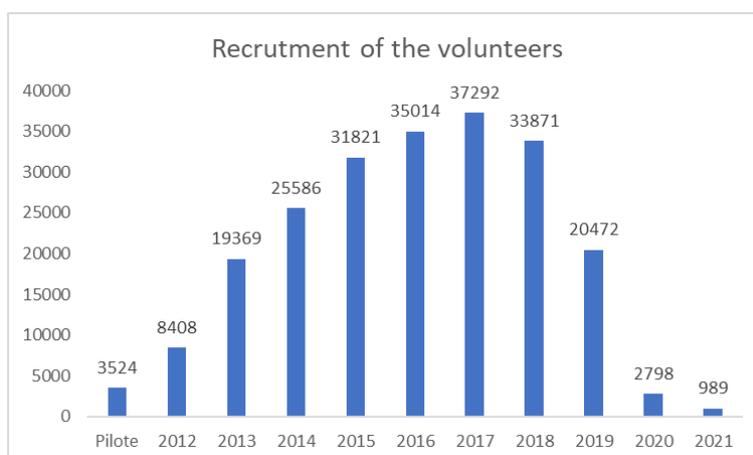
The *CONSTANCES* cohort is characterized by a broad coverage of health problems and determinants and an openness towards diversified users. *CONSTANCES* is an open epidemiological infrastructure, like large scientific instruments (such as a telescope, for example, or a genotyping laboratory equipped with sequencers), which are not built to answer a specific question, but are designed to help analyze a wide range of scientific problems, and are accessible to the specialized research and public health community.

2 METHODS: MAIN ELEMENTS OF THE PROTOCOL

The *CONSTANCES* cohort relies on two existing national systems on which it is largely based thanks to the partnership with the CNAM: (i) the Social Security Health Examination Clinics (HECs), present throughout the country, benefiting from a technical platform allowing the collection of biomedical data under particularly favorable conditions; (ii) the national administrative databases of the National Health Data System-SNDS managed by the CNAM and the National Old Age Insurance Fund, allowing access to health and socio-professional data.

2.1 SETTING UP AND FOLLOW-UP OF THE COHORT: OVERVIEW

CONSTANCES is a prospective epidemiological cohort. The sample that constitutes the cohort aims to be representative of the French population aged 18 to 69 years at inclusion. The total number of subjects is circa 220 000, and its structure is proportional to the population for sex, age, and social category. *CONSTANCES* was implemented in 2012, after a major pilot phase in 2009-2010. Recruitment of participants was mostly from 2012 to 2019, ending in 2021, when it had more than 220,000 volunteers.



The main steps in setting up and follow-up of the cohort are summarized here.

- Selection of eligible subjects: eligible subjects are drawn at random from a database managed by the CNAV, which includes all persons living in France.

- Invitation to participate: the persons selected receive a letter presenting the *CONSTANCES* project, an information leaflet on the HECs, and a reply coupon allowing them to give their agreement to participate in the cohort.
- Inclusion of the volunteers: subjects are given a health examination to collect health data and complete questionnaires. A consent form is proposed to the participants, which they must sign if they agree to participate.
- Active follow-up: a self-questionnaire (by mail or Internet) to be completed at the subjects' homes is proposed every year, and an invitation to come to the HEC every 4 years for a new health examination.
- Passive follow-up of socio-professional events and health data: the main socio-professional events and health data are regularly extracted by linkage to the hospital and claims administrative database (SNDS) and the CNAV database.

2.2 WHO IS IN THE COHORT

The cohort, composed of 54% women, constitutes a broadly diversified sample in terms of age and social characteristics.

Age class (January 2022)	All		Men		Women	
	N	%	N	%	N	%
[18-25]	1 466	0.7	612	0.7	854	0.8
[25-30]	7 268	3.6	2 901	3.1	4 367	4
[30-35]	16 226	8	7 258	7.7	8 968	8.2
[35-40]	19 976	9.8	9 127	9.7	10 849	9.9
[40-45]	21 938	10.8	9 973	10.6	11965	11
[45-50]	24 152	11.9	11 125	11.8	13 027	11.9
[50-55]	21 161	10.4	9 799	10.4	11 362	10.4
[55-60]	20 948	10.3	9 686	10.3	11 262	10.3
[60-65]	22 318	11	10371	11	11 947	10.9
[65-70]	22 460	11	10 860	11.5	11600	10.6
70 +	25 529	12.6	12 418	13.2	13 111	12

Distribution of age groups

Years of education	Total		Men		Women	
	N	%	N	%	N	%
< 8 years	5 697	2.8	3 011	3.2	2 686	2.4
[9-11] years	42 591	20.9	22 422	23.8	20 169	18.4
[12-13] years	32 605	16	14 348	15.2	18 257	16.7
[14-15] years	51 587	25.4	20 350	21.6	31 237	28.6
16 years	16 719	8.2	6353	6.7	10 366	9.5
17 years+ plus	48 303	23.7	24 740	26.3	23 563	21.6

Distribution of education levels

Size – Power

The large size of *CONSTANCES* makes it possible to answer multiple questions in a wide variety of

fields, by conducting epidemiological studies with good statistical power for sufficiently frequent diseases and/or risk factors. For example, among the participants in 2021, there were approximately 34,000 cases of cancer registered in the SNDS, of which more than 5,300 were incident during the follow-up of the cohort; for cardiovascular events, the figures are 16,800 and 6,700 respectively. These figures, which correspond to the situation in 2021, will increase with the length of follow-up of the cohort. Regarding the main risk factors, 28% of the participants are, according to the AUDIT score, in a situation of alcohol abuse, and 4.8% in a situation of dependence. While 42.5% of the volunteers have never smoked, 19.4% are current smokers and 38.1% ex-smokers. Twenty-three percent of men and 10% of women are or have been exposed to solvents in the workplace during their working career; 20% of men and 15% of women have or have had atypical working hours, including alternating shifts and night work. Thanks to these large numbers, many well-powered studies have already been carried out concerning very diverse aspects of health in relation to risk factors of various kinds.

Another problem affecting the power of a longitudinal study is that of lost to follow-up. Thanks to linkage with the SNDS and CNAV administrative databases, the number of "true" dropouts (people who ask to leave the cohort, registration errors, subjects leaving France permanently) is negligible. Active participation by annual questionnaires averages about 70% each year, and by 2021 only less than 0.5% had asked to leave the cohort.

2.3 MAIN DATA COLLECTED

The data to be collected systematically for all participants is intended to provide a corpus for describing and following over time the evolution of health status, general morbidity and mortality, socio-economic and occupational status, family and social environment and living environment, personal and environmental risk factors.

Socio-demographic characteristics, social situation

Employment status and activity, level of education, income level, marital status, household composition, socioeconomic status of parents and spouse, material living conditions, geographic origin of the subject and his/her parents.

Health data

Personal and family medical history, self-reported health scales (self-rated health, quality of life, mental health, specific scales: cardiovascular, musculoskeletal disorders, respiratory health), pathologies (list of reported diseases, diagnosis of long-term chronic diseases, hospitalizations, absence from work, disabilities, limitations, incapacities and traumas, use of care and management), date and medical cause of death.

In the HEC: clinical and paraclinical examination (weight, height, waist-to-hip ratio, blood pressure, heart rate, vision, hearing, spirometry, electrocardiogram), biological investigations (glycemic regulation, lipid balance, liver balance, creatinemia, blood count, urinary examination).

For people aged 45 and over, specific examinations are performed:

- Assessment of functional abilities: IALD scale (Instrumental Activities of Daily Living), ability to use new technologies.
- Cognitive functions: MMSE; Trail Making Test A - B; Wechsler Code; Semantic and Alphabetic Lexical Evocation; Grober & Busckhe RL-RI 16 memory test.
- Physical functioning: walking speed; Digital Finger Tapping Test; balance test (unipodal station); Hand Grip Test.

Lifestyle

Tobacco and alcohol consumption, diet and physical activity, cannabis use, sexual orientation.

Occupational factors

Occupational history, occupational exposure to chemical, physical and biological agents, postural, gestural and organizational constraints, work stress.

Specific health problems of women

Menopausal treatments, osteoporosis and osteoporotic fractures, sphincter and perineal disorders, mastopathies, endometriosis and chronic pelvic pain, infertility and delay in procreation, sex life and sexually transmitted diseases

Biobank

A bank of biological material is being set up for the health check-up. It is financed for the conservation of 26 aliquots of blood (serum, lithium heparin plasma, EDTA plasma and buffy coat) and urine for 60,000 subjects. It is managed by the Integrated Biobank of Luxembourg-IBBL.

2.4 DATA SOURCES

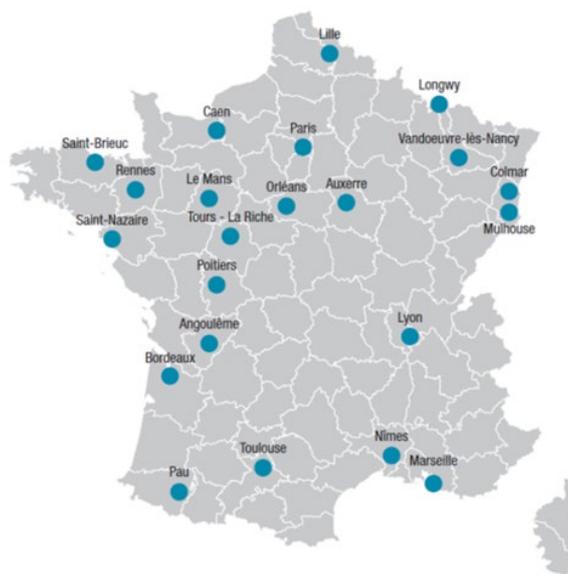
At inclusion and during follow-up, participants are subject to active and passive data collection.

Active collection

Questionnaires: at inclusion, several questionnaires are completed: Lifestyle and health, Women's health, Job history, Occupational exposures, past and current main diseases completed by the physicians after the clinical examination. During the follow-up, an annual questionnaire is sent to the participants, in paper form or online, according to the volunteers' choice.

Clinical and paraclinical examinations: the examinations are carried out at the time of inclusion and periodically during follow-up, in Social Security Health Examination Centers (HEC) throughout France.

There are currently 21 HECs participating in the project. These centers are large, have sufficient and motivated staff, a high-quality technical platform, and offer a geographical distribution allowing to represent the main French regions. The participating HECs are located in 20 different "départements".



CONSTANCES' recruitment centers (HECs)

Passive collection: linkage to national administrative databases

A so-called passive collection (because it does not involve any intervention by the subjects themselves) allows the extraction of data from national administrative databases

National Health Data System-SNDS: the hospital and claims database (SNDS) contains individual medicalized health data, structured and coded in a standardized way. The Health Insurance claims database includes details of all reimbursed services (consultations with health professionals, drugs and medical devices). The severe chronic disease database concerns all diseases exempted from co-payment, coded according to the International Classification of Diseases (ICD 10). The hospital database includes, for each hospital stay, the main diagnosis, the associated diagnosis(es), and procedures. Diagnoses are coded according to the ICD 10 and procedures according to the "*Classification Commune des Actes Médicaux*" (CCAM).

National Old Age Insurance Fund (CNAV): the CNAV databases concern socio-professional events. Since the CNAV's role is to ensure the right to retirement pension payment for any individual, it collects since the first job, social data from employers and from various organizations and schemes that manage social benefits: type of job and contract, salary, periods of activity/non-activity of individuals (unemployment, absence due to illness, maternity leave, etc.).

2.5 VOLUNTEER RECRUITMENT AND DATA COLLECTION AT INCLUSION

Volunteer recruitment took place over several years from 2012 to 2019, in annual waves according to the following procedures.

Constitution of the cohort

A random draw was made in the "*Répertoire national inter-régimes de l'assurance maladie*" (national data base managed by CNAV which includes all persons in France, who are registered by the social security, i.e. the whole population), among the eligible persons (18-69 years old residing in the "*départements*" (French administrative districts) where the *CONSTANCES* HECs are located). In order to obtain a sample close to the general population, the sampling scheme was stratified by age, sex, socioeconomic status and region of France and was based on unequal probabilities, to take into account the probabilities of participation according to age and socioeconomic status observed in surveys already carried out. Each subject drawn is given an individual non-identifying number "NConstances".

In France, each person has a unique national identifier that is used in the SNDS and CNAV databases to assign personal data to individuals. These identifiers were transmitted by the CNAV to the CNAM, which generated a new non-identifying individual number, different from the previous one, for linking to its databases.

Invitation to participate

An invitation to participate was sent to the eligible subjects, along with documents specifying the purposes and procedures of the cohort. Concerning information to individuals on the processing of their data, the documents specified that this study has been authorized by the French data protection authority, and that they have the right to access, modify, rectify and delete their personal data.

Inclusion of participants

People who volunteer to participate in the cohort received questionnaires to complete at home and went to a HEC, where they underwent clinical and paraclinical examinations, the procedures for which are set out in Standard Operating Procedures (SOPs) designed to ensure strict standardization. The examinations performed are as follows: biometry, vision, hearing, spirometry, electrocardiogram, blood pressure, functional cognitive and physical tests for those aged 45 and over. Quality control procedures are regularly implemented. The

health examination results data, identified by the individual non-identifying NConstances number, are sent to the *CONSTANCES* team; the completed questionnaires are sent to a service provider for data entry and then transmitted to the *CONSTANCES* team. Blood and urine are collected to set up a biobank managed by IBBL.

During the visit to the HEC, an information booklet on the *CONSTANCES* project was given to the participants and a consent form for the different data collections was proposed; the volunteers had to sign this informed consent to be part of *CONSTANCES*. A specific information booklet and consent form were also proposed for the biobank.

Confidentiality of identifying data: trusted third party

For contact with participants, a system including a "trusted third party", an organization independent of *CONSTANCES*, which manages the identifying information of the volunteers, has been set up. The trusted third party manages the personal data and addresses of the volunteers and establishes a correspondence table with the individual non-identifying NConstances number. It has no access to any other data about the participants. All exchanges with the participants (letters, emails...) are made through the trusted third party, the volunteers being identified in the *CONSTANCES* database only by their individual non-identifying number.

Passive data collection: linkage to national administrative databases

CNAV data: for each inclusion wave, the CNAV extracts from its databases the vital status and socio-professional data of participants and non-participants, which was sent to the *CONSTANCES* team. Regardless of their year of inclusion, the data of participants were extracted from their first job.

SNDS data: for each inclusion wave, the CNAM applied an algorithm to generate the number allowing access to the SNDS for each subject, and extracts from its information system the data of participants. Regardless of their year of inclusion, participant data have been extracted since 2007, the first year of SNDS data availability.

2.6 FOLLOW-UP

Active participation of the subjects in the follow-up

An annual self-questionnaire (by mail or Internet) is to be completed at home. Volunteers are also invited every 4 years to their HEC for a follow-up health examination. In order to maximize the participation rate in the follow-up, regular contact with the participants takes the form of a "Journal of the *CONSTANCES* cohort" presenting the results obtained, the associated projects, etc., sent regularly to the participants, a quarterly newsletter and a website.

Passive collection of socio-professional events and health data

Volunteers included in *CONSTANCES* are monitored for health data and socio-professional events by linkage to the national administrative databases from the CNAV and the SNDS. Each year, the CNAV and the CNAM send to the *CONSTANCES* team new socio-professional data, vital status and health data of participants.

Quality control

The self-administered questionnaires are subject to the usual controls: percentage of non-response, missing data, return time, etc.

For data collected during the inclusion visit to the HEC, a systematic and permanent quality control has been set up in order to evaluate the accuracy, reproducibility, concordance, internal and external validity of the data collected, and to study the factors of variability. Research technicians make regular site visits to verify compliance with standard operating procedures.

For data extracted from national databases, particular attention is paid to the validation of diagnoses extracted from SNDS, which are systematically checked.

Data storage

The data are associated for each subject with their permanent non-identifying number and integrated into the *CONSTANCES* database which is stored on the infrastructure of the "Secure Data Access Center" (CASD: www.casd.eu/), which offers all security guarantees.

Periodicity of data collection

The self-administered questionnaire is completed annually, which allows for close follow-up, the collection of a large amount of data without requiring too much effort from the volunteers, rapid response to set up new studies, and retention of participants, since a long delay between questionnaires is a factor in dropout. Some data are collected annually (health status and self-reported morbidity, life events and characteristics of place of residence, smoking, alcohol, etc.); others are collected at longer intervals according to a pre-established schedule (health scales, questionnaires specific to a health domain or risk factors such as diet).

The follow-up in the national administrative databases is permanent, since they record events essentially continuously.

Participants are invited every 4 years to the HEC for a follow-up health examination.

2.7 SELECTION EFFECTS AND REPRESENTATIVENESS

The objectives of *CONSTANCES* are to be a support platform for research on the determinants of health, but also to provide descriptive data on the health and risk factors of the French population.

The fact that participation in the cohort is voluntary introduces selection effects, as participants differ from non-participants in many aspects. The problem of selection bias is very different for analytical and descriptive purposes. In a cohort in which the inclusion procedures were the same for all subjects, the exposure-disease relationship is not a priori different between those included and those not included. When it comes to analytical studies, the selection procedures for *CONSTANCES* participants will therefore not generate a priori any bias, or only minimal bias. The problem of attrition during follow-up, however, may cause significant bias if the probability of no longer being followed up differs for exposed and unexposed individuals and/or for those who have or have not become ill, which is often the case.

For the **descriptive study of the frequency of health problems and risk factors**, the parameters of interest must be estimated on a representative sample of the target population. It has been verified that the population structure of all the "*départements*" where the *CONSTANCES* HECs are located is practically identical to that of the whole of France for the main demographic and socio-professional characteristics; the results of *CONSTANCES* can therefore be generalized to the French population. To obtain a representative sample of the target population and to minimize biases related to selection effects at inclusion and during follow-up, the following methods were implemented.

Random drawing of eligible participants

The draw was made from a sampling frame made up of all the members of the social security system in the "*départements*" where the *CONSTANCES* HECs are located, aged 18 to 69 years. The sampling scheme was stratified by age, sex, socioeconomic status and region of France with unequal probabilities according to age, sex and SES, over-representing the categories that participate less in health surveys.

Adjustment for non-participation

A random sample of non-participants, twice as large as the sample of participants, was constituted as inclusions were made to serve as a "reference cohort". For these subjects, we have socio-demographic characteristics, thanks to the CNAV files, as well as a lot of information on health and health care use, thanks to the SNDS data. Thus, for the participants and the sample of non-

participants, we have the same data from the SNDS and CNAV files. These "auxiliary" data make it possible to estimate the probabilities of participation in *CONSTANCES* using prediction models; the inverse of the probability of participation then provides the adjustment coefficient for each participant. It is thus possible to carry out weightings taking into account the effects of selection at inclusion; the richness of the socio-professional and health data used allows very effective adjustments.

For the calculation of prevalences, the cohort is calibrated to the margins of the reference population, using the random sample of non-participants to calculate the relevant margins that incorporate sociodemographic, economic and health-related data. The quality of the estimates is thus significantly improved by calculating margins that are specifically related to the health purpose of the cohort. The development of these methods has enabled *CONSTANCES* to receive the Label of General Interest and Statistical Quality from the National Council for Statistical Information (CNIS: www.cnis.fr/).

Adjustment for attrition

It can be considered that, apart from recording errors, none of the persons included in *CONSTANCES* will be definitively lost to follow-up, since the participants are passively followed through the SNDS and CNAV administrative databases. There will, however, be attrition due to failure to return the annual self-questionnaire or not to return to the HEC for new health examinations. Since the data collected at the time of inclusion in *CONSTANCES* as well as the data from the SNDS and CNAV files corresponding to the year of inclusion are available for all participants, adjustment coefficients for attrition can be calculated using a method similar to that for calculating adjustment coefficients for non-participation. The passive and longitudinal follow-up of participants in the SNDS and CNAV databases, whether or not they drop out of the cohort because they do not return a self-administered questionnaire, makes it possible to regularly update the attrition adjustment coefficients.

2.8 REGULATORY ISSUES

All the procedures described have been the subject of regulatory formalities with the authorities concerned: Commission nationale de l'informatique et des libertés (data protection authority in France), Inserm's Institutional Qualification Committee, Committee for the Protection of Individuals. All procedures used for the creation, management and use of the cohort are compliant with the General Data Protection Regulation (GDPR).

2.9 REFERENCES

The main aspects of the *CONSTANCES* protocol have been published.

Zins M, Goldberg M and the *CONSTANCES* team. The French *CONSTANCES* population-based cohort: design, inclusion and follow-up. *Eur J Epidemiol.* 2015;30(12):1317-28. doi: 10.1007/s10654-015-0096-4.

Goldberg M, Carton M, Descatha A, Leclerc A, Roquelaure Y, Santin G, Zins M; *CONSTANCES* team. *CONSTANCES*: a general prospective population-based cohort for occupational and environmental epidemiology – Cohort profile. *Occup Environ Med.* 2017;74(1):66-71. doi: 10.1136/oemed-2016-103678.

Ruiz F, Goldberg M, Lemonnier S, Ozguler A, Boos E, Brigand A, Giraud V, Perez T, Roche N, Zins M. High quality standards for a large-scale prospective population-based observational cohort: *CONSTANCES*. *BMC Public Health.* 2016;16(1):877-886. doi: 10.1186/s12889-016-3439-5.

Henny J, Nadif R, Le Got S, Lemonnier S, Ozguler A, Ruiz F, Beaumont K, Brault D, Sandt E, Goldberg M, Zins M. The *CONSTANCES* Cohort Biobank: An Open Tool for Research in Epidemiology and Prevention of Diseases. *Front. Public Health* 2020; 8:605133. doi: 10.3389/fpubh.2020.605133.

2.10 PUBLICATIONS FROM CONSTANCES

Publications resulting from research conducted in the cohort can be consulted on the Constances website: www.constances.fr/espace-scientifique/productions-scientifiques.php#publications-comite-lecture.

3 ACCESS TO THE *CONSTANCES* COHORT

All teams, French or foreign, public or private, can ask to use the *CONSTANCES* cohort for their research work. It can be to use available data as well as collected biological materials, and/or to collect additional data from volunteers for a specific purpose. Applicants must submit an application detailing the scientific and methodological aspects of the project, its possible implications regarding confidentiality, ethics and deontology, as well as its financing and a provisional timetable for its completion. Applications are reviewed by the International Scientific Council of *CONSTANCES*.

A Charter for access to the *CONSTANCES* cohort specifies the terms and conditions for the access to the infrastructure: terms and conditions for the selection of projects, conditions for the exploitation of available data, terms and conditions for the collection, exploitation and access related to data collected by complementary investigations, financing of projects, scientific responsibilities, data protection, dissemination of research results).