BMJ Open

The Etiosarc Study : environmental aetiology of sarcomas from a French multicentric population-based case-control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030013
Article Type:	Protocol
Date Submitted by the Author:	22-Feb-2019
Complete List of Authors:	Lacourt, Aude; Université de Bordeaux, Inserm U1219 - EPICENE team Amadéo, Brice; Université de Bordeaux, Inserm U1219 - EPICENE team; Université de Bordeaux, Inserm CIC 1401, Registre des cancers de la Gironde Gramond, Céline; Université de Bordeaux, Inserm U1219 - EPICENE team Marrer, Emilie; Groupe Hospitalier de la Région de Mulhouse et Sud Alsace, Registre des cancers du Haut-Rhin Plouvier, Sandrine; Registre des cancers de Lille et de sa Région Baldi, Isabelle; Université de Bordeaux, Inserm U1219 - EPICENE team; CHU de Bordeaux, Pôle de Santé Publique, Service de Médecine du Travail et Pathologies professionnelles Blay, Jean-Yves; Université Claude Bernard Lyon 1, Department of Medical Oncology, Centre Léon Berard Coindre, Jean-Michel; Institute Bergonie, Department of Biopathology de Pinieux, Gonzague; CHRU Tours, service d'anatomopathologie Gouin, François; CHU Nantes, Inserm UI957 Italiano, Antoine; Institute Bergonie, Department of Medical Oncology Le Cesne, Axel; Institut Gustave Roussy, Medical Oncology Department Le Loarer, François; Institut Bergonié, Department of Biopathology Monnereau , Alain; Université de Bordeaux, Inserm U1219 - EPICENE team; Institute Bergonie, Gironde registry of haematological malignancies Pellegrin, Isabelle; CHU de Bordeaux, Service d'Immunologie-Immunogénétique, Pôle de Biologie et Pathologie Penel, Nicolas; Ctr Oscar Lambret, Medical Oncology Department Ray-Coquard, Isabelle; Université Claude Bernard Lyon 1, Department of Medical Oncology, Centre Léon Berard Toulmonde, Maud; Institute Bergonie, Department of Medical Oncology Ducimetière, Françoise ; Centre Léon Bérard, Réseau de référence sarcome Mathoulin-Pélissier , S; Université de Bordeaux, Inserm U1219 - EPICENE team; Institute Bergonie, Clinical and Epidemiological Research Unit, Inserm CIC 1401
Keywords:	cancer, Sarcoma < ONCOLOGY, case-control study, environmental and occupational exposure, aetiology

The Etiosarc Study : environmental aetiology of sarcomas from a French multicentric population-based case-control study

Lacourt A[1], Amadéo B[1,2], Gramond C[1], Marrer E [3], Plouvier S[4], Baldi I[1,5], Blay JY[6], Coindre JM[7], De Pinieux G[8], Gouin F[9], Italiano A[10],Le Cesne A[11], Le Loarer F[7], Monnereau A[1,12], Pellegrin I[13], Penel N[14], Ray-Coquard I[6], Toulmonde M[10], Ducimetière F[15], Mathoulin-Pélissier S[1,16] for the Etiosarc study group

1 Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, Epicene team, UMR 1219, F-33000 Bordeaux, France

2 Univ Bordeaux, Inserm CIC1401, Registre des cancers de la Gironde, Bordeaux, France

3 Registre des cancers du Haut-Rhin, Groupe hospitalier de Mulhouse et Sud Alsace (GHRMSA), Mulhouse, France

4 Registre Général des Cancers de Lille et de sa Région, Lille, France

5 CHU de Bordeaux, Pôle de Santé Publique, Service de Médecine du Travail et Pathologies professionnelles, Bordeaux, France

6 Université Claude Bernard Lyon I, Department of Medical Oncology, Centre Leon Berard, Lyon, France

7 Department of Biopathology, Institute Bergonié, Bordeaux, France

8 Department of Pathology, CHU de Tours, Tours, France

9 CHU Hôtel-Dieu, Nantes, France; INSERM UI957, Nantes, France

10 Department of Medical Oncology, Institute Bergonié, Bordeaux, France

11 Department of Medical Oncology & Surgery, Institut Gustave Roussy, Villejuif, France

12 Gironde registry of haematological malignancies, Institut Bergonié, F-33000 Bordeaux, France

13 Service d'Immunologie-Immunogénétique, Pôle de Biologie et Pathologie, Groupe Hospitalier Pellegrin, CHU Bordeaux, Bordeaux, France

14 Medical Oncology Department, Centre Oscar Lambret, Lille, France

15 Réseau de référence sarcome, Centre Léon Bérard, Lyon, France

16 Clinical and Epidemiological Research Unit, INSERM CIC1401, Institut Bergonié, Comprehensive Cancer Center, F-33000 Bordeaux, France

Etiosarc study group : Lacourt A, Amadéo B, Gramond C, Marrer E, Plouvier S, Baldi I, Bara S, Bazille C, Blay JY, Bompas E, Chaigneau L, Chateau MC, Coindre JM, Coureau G, Cupissol D, D'Almeida T, Defossez G, Delafosse P, Delcambre Lair C, De Pinieux G, Di Marco A, Fabre T, Fiorenza F, Ghnassia JP, Gouin F, Guizard AV, Italiano A, Kurtz JE, Lebrun-Ly, Le Cesne A, Le Loarer F, Le Nail LR, Maynou C, Missenard G, Molinié F, Monnereau A, Moreau A, Penel N, Ranchère-Vince D, Ray-Coquard I, Robin YM, Terrier P, Toulmonde M, Tretarre B, Velten M, Woronoff AS, Ducimetière F, Mathoulin-Pélissier S.

CORRESPONDING AUTHOR

Aude Lacourt
Equipe EPICENE Epidemiology of cancer and environmental exposures
Bordeaux Population Health Center - Inserm U1219
Université Bordeaux - ISPED
146 rue Leo Saignat
33076 Bordeaux Cedex - France
Tel:+33 5 57 57 16 23 - Fax: +33 5 57 57 47 33
aude.lacourt@inserm.fr

ABSTRACT

Introduction: Sarcomas are rare tumors of connective tissue. The exact overall incidence of sarcomas is unknown due to diagnostic difficulties and the various histological subtypes (over 80 subtypes). However, the apparent increasing incidence of sarcomas suggests environmental causes such as pesticides. Except for some specific factors (i.e. ionizing radiation, vinyl chloride, dioxin, and genetic predispositions) the scientific knowledge on the aetiology of sarcomas is sparse and inconsistent. France is a particularly appropriate country to set up a study investigating the causes of sarcoma occurrence due to the French organization in treatment and care of sarcoma patients, which is highly structured and revolved around national expert networks. The main objective of the ETIOSARC project is to study the role of lifestyle, environmental and occupational factors in the occurrence of sarcomas among adults from a multicentric population-based case-control study.

Methods and analysis: Cases will be all incident cases (older than 18 years old) identified in 15 districts of France covered by a general population-based cancer registry and/or a reference center in sarcoma's patient care over a three-year period with an inclusion start date ranging from the 1st February 2019 to the 1st January 2020 and histologically confirmed by a second review of the diagnosis. Two controls will be individually-matched by sex, age (5-years group), and districts of residence and randomly selected from electoral rolls. A standardized questionnaire will be administered by a trained interviewer in order to gather information about occupational and residential history, demographic and socioeconomic characteristics and lifestyle factors. At the end of the interview, a saliva sample will be systematically proposed.

This study will permit to validate or identify already suspected risk factors for sarcomas such as phenoxyherbicides, chlorophenol and to generate new hypothesis to increase our understanding about the genetic and environmental contributions in the carcinogenicity process.

Count words: 4480

Keywords: cancer, sarcoma, case-control study, environmental and occupational exposure, aetiology

Ethics and dissemination: This study received National French Ethic committee approval (identification number 2018-A00519-46) and French Data Protection Authority (CNIL) approval (identification number 918171). Results of this study will be published in international peer reviewed journals.

Technical appendix, statistical code, and dataset will be available in the Dryad repository when collection data is completed.

Article Summary

Strengths and limitations of this study

- All newly diagnosed cases will be ascertained through a systematic review of the diagnostic by an expert pathologist within French sarcomas networks.
- Inclusion of more than 2000 population-based cases which will allow to perform analysis by histological subtypes.
- Collection of various exposure data using face-to face administered standardized questionnaire by trained interviewers. Additionally, saliva samples will be collected
- Possible low participation rate of population controls.
- Retrospective collection of data and over a long period of time.

INTRODUCTION

Sarcomas are a heterogeneous group of rare malignant tumors of connective tissues. Besides multiple and complex histology (approximatively over 80 subtypes), these tumors can occur in almost any anatomic site. It is usual to distinguish bone sarcomas (osteosarcomas and chondrosarcomas) from soft-tissue sarcomas (muscles, joints, fat, nerves, skin tissues and blood vessels) and visceral sarcomas. Even though it accounts for less than 1% of adult cancers, sarcoma is one of the most frequent cancer types in young adults [1]. Based on statistics of the Surveillance Epidemiology and End Results (SEER) program from 2008-2012, median age at diagnosis for soft-tissue sarcomas and bones sarcomas were 59 and 42 years of age, respectively and median age at death was 65 years old for soft-tissue sarcomas and 59 years old for bones sarcomas [2].

At the genetic level, sarcomas can be split into two large categories based on the rearrangement level of the genome. On the one side sarcomas with very simple genetics (50% of all sarcomas) based on point mutation (gastrointestinal stromal tumor (GIST), desmoïd tumor) or a specific and recurrent translocation (Ewing, Synovial sarcoma, Myxoïd liposarcoma, …); and on the other side, sarcomas with a very complex genetics (50% of all sarcomas, leiomyosarcoma, angiosarcoma, undifferentiated pleomorphic sarcoma,…) [3].

The incidence of sarcomas has been very imperfectly estimated due to both diagnostic confusion with carcinomas of the same organ and the variety of localization of these tumors. Thirty percent of sarcomas are misclassified at initial diagnosis [4]. Cancer incidence is often reported by site possibly leading to an under-estimation of incidence for some subtypes. In France, the world age-standardized incidence rates of overall sarcomas was estimated at 4.8 per 100,000 inhabitants per year [5] and soft-tissue sarcoma's incidence, which account for more than 50% of sarcomas cases was estimated at 3.3 per 100,000 inhabitants per year [6]. While some authors did not confirm an incidence increase over last years [7], others reported a statistically significant raise of the incidence of sarcomas [8 9] leading to the hypotheses of

an implication of environmental factors in the aetiology of this cancer. However, the role of diagnostic and reporting practice cannot be ruled out.

The various histological subtypes, various anatomical sites as well as the rarity of the disease make the aetiology of this cancer difficult to study [1]. Some studies have attempted to investigate the role of some environmental factors; however, results are often inconclusive or inconsistent from one study to another. Thus, to date, it is impossible to draw strong conclusions on the aetiology of sarcomas from existing studies. Indeed, there are some methodological considerations to explain discrepancies between studies: sample sizes are often limited leading to insufficient power to detect small but relevant increases in risk. As a consequence, sarcomas are studied as a single outcome and not by histological sub-types; at most they are segregated into bone sarcomas and soft-tissue sarcomas. Such analyses support the strong hypothesis that the same aetiology is shared between each sub-type and each site. The inclusion periods of cases are usually old and ascertainment of cases may be incorrect. Indeed, a second expert review of diagnosis is essential to correctly classify sarcoma's tumors [10]. Diagnostic procedures have been refined in expert centers during the last several years with the inclusion of new genetic and molecular data leading to a better understanding and definition of tumors. As a consequence, a new World Health Organization classification of soft-tissue tumors was published in 2013 [11].

From IARC evaluation, the strongest evidence for environmental cause is for ionizing radiations (including radiotherapy) for both soft tissue and bone sarcomas that are classified in group 1 as carcinogenic to humans [12]. Convincing evidence also exists for linking vinyl chloride to the occurrence of a specific type of sarcoma (liver angiosarcoma). This conclusion arises from consistent observations in vinyl chloride industry that conducted IARC to classify this agent also in group 1 as carcinogenic in humans [13]. The dioxin 2,3,7,8-Tetrachlorodibenzo-p-dioxin (2,3,7,8-TCDD), released to the environment during the combustion of fossil fuels and wood, and waste incineration has also been classified carcinogenic by the IARC, with evidence for soft-tissue sarcoma [13].

Because epidemiological studies have consistently demonstrated higher mortality and incidence rates from soft-tissue sarcomas in farmers, specific attention has been paid to the potential role of pesticide exposure in the occurrence of these tumors [14]. Several case-control studies assessed the relationship between herbicide and chlorophenol with soft-tissue sarcoma risk but results varied from no association to a strong relationship for exposed populations [15-20]. Discrepancies between studies may be due in part to the low statistical power due to small sample sizes but most importantly to difficulties in exposure assessment. Indeed, disentangling the various contributions of chlorophenols, herbicides and dioxins is complicated because chlorophenols are used in the production of herbicides and dioxins are also a contaminant in herbicides production.

Some studies have also suggested that other pesticides than phenoxys could play a role in the occurrence of sarcomas. Increased risk of soft-tissue sarcoma have been observed in a cohort of workers in an organochlorine production plant [21], in a population living in the vicinity of an organochlorinated-compounds factory in Spain [22], and also in a case-control study in Canada exploring some specific pesticides [23]. This last study found significant associations with two insecticides: Aldrin (OR=3.71, 1.00-13.76) and Diazinon (OR=3.31, 1.78-6.23) and a trend with formaldehyde (OR=2.07, 0.94-4.56). A case-control study in Kansas failed to find associations with pesticides used on crops, including herbicides [24] but found an increased risk with the use of insecticides to animals [25], higher for farmers who mixed or applied themselves and for those who did not use any protective equipment, also more pronounced for so-called “fibrous” and “myomatous” sarcomas. A European study on risk of adult bone sarcomas also found an association with pesticides that was similar for insecticides and herbicides [26].

Other occupational exposures have also been explored and, besides farming, some industries have been associated in some studies with elevated risks of soft-tissue sarcomas: gardeners, meat packers, sawmill workers, machinists, ground maintenance workers [27-29]. Exploring specific chemicals or agents, positive associations have been found with wood

dust [28], radium [27] and 1-3 butadiene [30] for soft-tissue sarcoma. A multicenter case-control study in seven European countries specifically focused on bone sarcomas found increased risks among blacksmiths, toolmakers, machine-tool operators and construction workers [26].

Besides occupational factors, few studies investigated other environmental factors related to lifestyle, including, the potential role of female hormones, tobacco smoke, alcohol, nutrition, fluoride in drinking water, and body mass index. These studies reported inconclusive or inconsistent results with the exception of viruses like the human immunodeficiency virus (HIV) and human herpes virus 8 (HHV8) for Kaposi's sarcoma [1].

Finally, while there is increased evidence regarding genetic determinants of sarcoma risk [31], and growing evidence that gene × environment (G × E) interactions are determinants of development and progression of complex disease [32], to date, there is no data related to such G × E interactions with regards to sarcoma risk. However, studying G × E interactions may help to identify susceptible groups of individuals which is essential to better target prevention programs or to develop precision medicine [33].

Several hypotheses have been studied regarding the aetiology of sarcomas. Nevertheless, the methodological limitations such as diagnosis certification, small sample sizes, exposure assessment methodology, and the analysis strategies that do not distinguish the various histological subtypes making the hypothesis that the aetiology is homogeneous across all histological subtypes prevent any definite conclusions. There is a clearly identified need to further investigate the aetiology of sarcomas with improved study designs including a systematic centralized diagnosis review by a second expert, increased sample sizes, more refined exposure assessment methods and also a biological component to increase our understanding of biologic mechanisms of carcinogenesis and to study the interaction between the genetic and environment component.

France is a particularly appropriate country to set up such study due to the French organization in treatment and care of sarcoma patients around three national networks

labelled by the French National Cancer Institute (INCa): the network of reference for soft-tissue sarcoma pathology (RRePS: https://rreps.sarcomabcb.org/home.htm), the network of reference for bone sarcoma pathology (ResOs: https://resos.sarcomabcb.org/home.htm) and the clinical sarcoma network (NetSarc: https://netsarc.sarcomabcb.org/home.htm). The objectives of RRePS and ResOs are to ensure systematic and free secondary reviews for all new diagnoses of soft tissue and visceral sarcomas, gastro-intestinal stromal tumors (GIST), desmoid tumors and bone sarcomas across the whole of France, and to facilitate access to molecular biology analyses, collection of samples for biological resource centers, to participate in research and clinical trials, to draft good clinical practice guidelines for professionals and information documents for patients, and to organize continuing education and information for patients [34]. The NetSarc clinical network for sarcoma is the clinical network of the French Sarcoma Group dedicated to clinical patient care. NetSarc is managed by three sites (Centre Leon Berard in Lyon, Institut Bergonie in Bordeaux and Institut Gustave Roussy in Villejuif) working closely with 25 expert regional centers, ensuring good coverage of the whole French territory [35]. Besides, the French population-based cancer registries are organized in a collaborative network named Francim. The main objectives of this network are to coordinate the 14 general cancer registries and 11 specialized cancer registries that exhaustively register all newly diagnosed and confirmed cancer cases according to international procedures, to harmonize cases registration and data quality, to provide epidemiological indicators (incidence, survival, prevalence) and coordinate epidemiological and surveillance research on cancer.

The main objective of the ETIOSARC study is to assess the role of lifestyle, environmental and occupational factors in the occurrence of sarcomas among adults from a multicenter population-based case-control study.

Specific objectives are:

- To identify environmental risk factors for sarcomas as a whole and for the most frequent subtypes;

- To investigate the interactions between gene polymorphisms and environmental exposures in sarcoma susceptibility;
- To assess whether some specific genetic characteristics of sarcoma' tumors are associated with environmental exposures;

We will also explore the feasibility of classifying sarcomas by genetics types (simple vs complex genomic profile) instead of by histological subtypes as part of the objective of identifying environmental risk factors.

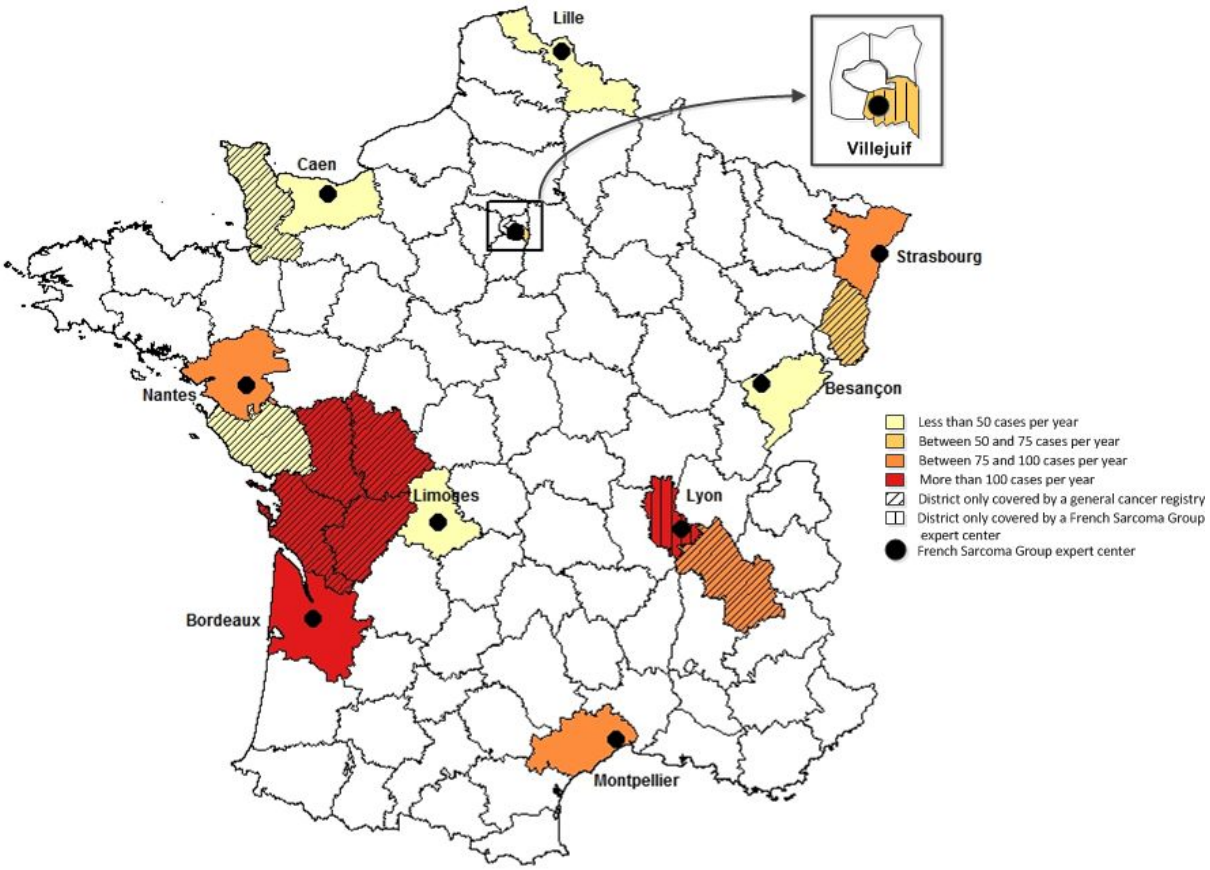
METHODS AND ANALYSIS

Study design

The ETIOSARC study is a prospective multicenter population-based case-control study. This study is restricted to French geographical areas (further called districts) that meet at least one of these four criteria:

- Criteria 1: Districts covered by both a general cancer registry and a French Sarcoma Group (GSF-GETO) expert center from the sarcoma reference network;
- Criteria 2: Districts including a coordinator center of any of the RRePS / NetSarc networks;
- Criteria 3: Districts covered by a general cancer registry that is expected to register more than 50 cases per year;
- Criteria 4: Districts adjacent to districts meeting criteria 1 and covered by a general cancer registry.

In total, 15 districts meet one of the four criteria, which represent 17,813,937 inhabitants, approximatively 27% of the French population (2019 estimations from the national institute of statistics and economic studies, Insee). Eight districts meet criteria 1 (Gironde, Hérault, Haute-Vienne, Loire Atlantique, Calvados, metropolitan area of Lille, Bas-Rhin and Doubs), two districts meet criteria 2 (Rhône and Val-de-Marne); three districts meet criteria 3 (Isère, Haut-Rhin and Poitou-Charentes) and two districts meet criteria 4 (Manche, Vendée).



< FIGURE 1: districts covered by the Etiosarc study >

Since sarcomas are rare tumors, this study has to be multicentric. Indeed, from general cancer registry data and RRePS overall incidence data, considering a patients' response rate of 70%, and considering that approximatively 90% of newly diagnosed cases will benefit from a systematic secondary review of diagnosis, it is expected to include 718 incident cases per year from these 15 districts, corresponding to a total sample size of 2 154 cases from a 3 years-recruitment.

< Table 1: Total number of expected cases includes in the Etiosarc study over a three years recruitment >

	Expected number of new sarcoma cases / year *	Expected number of confirmed cases / year†	Expected number of interviewed cases /year‡	Expected total number of included cases¶
Criteria 1: Districts covered by both a general cancer registry and a French Sarcoma Group (GSF-GETO) expert center from the sarcoma reference network				
Gironde	120	108	76	228
Hérault	86	77	54	162
Haute-Vienne	23	21	15	45
Calvados	50	45	31	93
Loire-Atlantique	98	88	62	186
Metropolitan area of Lille	48	43	30	90
Bas-Rhin	90	81	57	171
Doubs	42	38	27	81
Criteria 2: Districts including a coordinator center of any of the RRePS / NetSarc networks				
Rhône	116	104	73	219
Val de Marne	68	61	43	129
Criteria 3: Districts covered by a general cancer registry that is expected to register more than 50 cases per year				
Isère	89	80	56	168
Haut-Rhin	56	50	35	105
Poitou-Charentes	167	150	105	315
Criteria 4: Districts adjacent to districts meeting criteria 1 and covered by a general cancer registry				
Vendée	51	46	32	96
Manche	34	31	22	66
Total	**1 138**	**1 023**	**718**	**2 154**

* Estimation from general cancer registry data (except for Rhône et Val de Marne, estimation from RRePS data).

† Estimation with a systematic secondary review a diagnostic confirmation for 90% of cases.

‡ Estimation with a response rate of 70%.

¶ Over a three years study.

Study population

Cases definition and recruitment modalities

Cases are defined as all incident patients with a diagnosis of primary sarcoma and histologically confirmed by an expert pathologist of the RRePS or ResOs networks in the 15 districts of France participating to this study.

Inclusion criteria are:

- Patients diagnosed in the previous 6 months from identification with a primary and histologically confirmed malignant sarcoma including soft-tissue, visceral and bone sarcomas as defined by the WHO classification of bones and soft tissue sarcoma, fourth edition, 2013 [36];
- Diagnosed over a three-year period with an inclusion start date ranging from the 1st February 2019 to the 1st January 2020 depending on the districts;
- Living in one of the 15 districts participating to the study at the time of diagnosis;
- At least 18 years old at diagnosis;
- Agreed to participate to the study with a signed informed consent;

Non-inclusion criteria are:

- Patients with a known genetic predisposition to sarcoma such as Li-Fraumeni syndrome, Retinoblastoma syndrome, neurofibromatosis;
- Kaposi's' sarcoma;
- Protected adults' patients (aged of at least 18 years old) under guardianship by court order.

Patients will be recruited by specifically trained clinical research associate (CRA). In districts where a general cancer registry is active, the process of identifying incident sarcoma patients is well defined, which warrants the efficiency and exhaustiveness of the recruitment.

However, poor survival for some cancer cases will not allow to rely on routine inclusion procedure and the registries will implement a rapid case ascertainment procedure to minimize the delay between diagnosis and enrollment and interview. Cases will be identified from pathology laboratories and multidisciplinary sarcoma tumor board (including sarcoma, gynecological, digestive, skin and bone tumor board). CRA will regularly contact laboratories and hospitals (within a three months window) to identify newly diagnosed sarcoma cases and to collect associated reports. The CRA will retrieve the clinical file in order to check inclusion and non inclusion criteria and to collect names and addresses of eligible patients and their physicians. Then, the CRA will contact the case's physicians in order to gather their clinical advice to include their patients in the study. In case of agreement, the CRA will contact the patients and ask them to participate. In case of oral agreement, he/she will arrange an interview to collect written consent and administer detailed questionnaire.

In Rhône and Val de Marne where no general cancer registry exists, the process of identifying incident sarcoma patients will largely rely on the existence of a sarcoma reference network coordination center to facilitate the identification of new cases.

Regardless of the district, the diagnosis of all included cases will be systematically ascertained by an expert pathologist within the RRePS and ResOs networks following a standard procedure.

On a regular basis (at least once a month), the list of incident cases included in the RRePS/ResOS/NetSarc networks will be extracted from the shared databases. The list of cases identified by registries will be merged with the list of cases included in the RRePS/ResOS/NetSarc networks by the pathology sample reference number in order to collect diagnosis confirmation and to identify new but missed incident cases.

Controls selection

Two control subjects per case will be randomly selected from the French general population using electoral rolls and individually matched to case by age (by 5-year age group), sex and residential area (French department).

Inclusion criteria are:

- Subjects registered within the electoral rolls;
- At least 18 years old at interview;
- Living in one of the 15 districts participating to the study at the time of interview;
- Agreed to participate to the study with a signed informed consent;

Non-inclusion criteria are:

- Subject previously diagnosed with a primary and histologically confirmed malignant sarcoma including soft-tissue, visceral and bone sarcomas as defined by the WHO classification of bones and soft tissue sarcoma, fourth edition, 2013;
- Protected adults' patients (aged of at least 18 years old) under guardianship by court order.

The selection of controls and the recruitment of cases will take place simultaneously. Every time a case will be diagnosed and identified, two controls will be randomly selected at that time from the electoral rolls of the same district as the case district of residence. For each case, a list of 20 potential controls with the same age (within a 5-year age group) and sex as the cases will be constituted and randomly selected in order to account for potential refusal. Calls to contact potential controls will be centralized and will be made by an investigator specifically trained to better understand the objectives of the study. First, this investigator will try to retrieve the phone number of the first potential control on the list and will contact him/her in order to gather his/her agreement to participate in this study. After five calls made at different schedules (e.g. in the evening on weekdays), if the potential control couldn't be reached, the investigator will send an information letter containing a reply coupon indicating

phone number and schedule at which the person could be reach. If the reply coupon is not returned, or after 10 calls made at the phone number and the schedule indicated on the reply coupon, the first potential control will be considered as unreachable and then, the same procedure will be applied for the second potential control on the list and so on until two controls per case agreed to participate to the study.

For each control agreeing to participate, an appointment will be made to collect written consent and to administer detailed questionnaire by a CRA.

All participants will give their informed written consent to participation, in line with French ethical guidelines.

Data collection

Data will be collected in a three-step procedure and from subjects themselves: i) the interviewer will contact subjects by phone and ask them to participate in the study; ii) if the subject has agreed to participate, an appointment will be made and a consent form as well as a self-administered questionnaire will be sent by mail to each subject. This self-questionnaire will permit to gather the complete occupational history (for each job held for at least 6 months) and residential history (for each place occupied for at least 1 year); iii) during a face-to-face interview, the trained interviewer will check (complete if necessary) and supplement the self-administered questionnaire by a specific questionnaire with questions about demographic and socioeconomic characteristics, occupations of spouse and parents, leisure-time activities, reproduction, medical history, family history of cancer, diet, lifestyle factors such as tobacco smoking and alcohol consumption. The specific questionnaire will also collect additional occupational and residential information such as work tasks, work places, materials handled for each job held for at least 6 months and description of the environment of each residence places.

At the end of the interview, subjects (both cases and controls) will be invited to provide a saliva sample in order to obtain germline DNA.

In case of refusal to participate and if the subject agreed, data on the last occupied job and educational level will be collected in order to assess the potential selection bias that might occur due to the specific profile of non-respondent subjects.

To ensure consistency in data collection, all CRA will be trained in the completion of the questionnaire and will detain a field guide of completion. Additionally, phone meetings involving all CRA will take place each month in order to deal with recurring problems in the completion of the questionnaire and to ensure homogeneity in data collection between all CRA. Moreover, these meetings will help maintain CRAs' motivation and level of training.

The completeness of questionnaires and the quality of interviews will be routinely checked. All questionnaires will be scanned in order to facilitate storage and possible return to the questionnaires to allow quality checks.

Biological sampling and storage

Each participant will be asked to provide a salivary sample in order to obtain germline DNA. Salivary samples will be collected by the subjects themselves under instructions from the CRA, using Genefix™ saliva DNA collection kit. After collection, samples will be sent at the Biological Resources Center of the Bordeaux hospital university center "Bordeaux Biothèque Santé" for storage (NFS-96900 certification, BBMRI-ERIC ID: FR_BB-0033-00094).

Determination of the sample size

The main objective of this study is to examine the association between environmental exposures (including general environment, occupational environment and lifestyle) and risk of sarcoma occurrence. Since our definition of environmental exposures is very broad, we have based our sample size calculation on various scenarios of exposure prevalence, from

5% (relevant for domestic, environmental but also some occupational exposures such as farmer in the general population) to 20% (relevant for some occupational exposures such as fibers). For a total sample size of 2000 cases and a 1:2 individually-matched design, considering a statistical power of 80% at a significance level of 5%, the minimum detectable odds ratio will be 1.39 and 1.21 under an exposure prevalence of 5% and 20%, respectively. Considering sub-types analyses, the four main histological types are GIST, liposarcoma, leiomyosarcoma and unclassified sarcoma, which account for 18%, 15%, 11% and 16% of sarcomas, respectively. Considering a sample size of 300 cases, the minimum detectable odds ratio will be 2.21 and 1.61 under an exposure prevalence of 5% and 20%, respectively. Typically in environmental epidemiology, the relative increases in disease risks due to environmental exposures are usually low around 1.5, thus, it is essential to recruit a minimum of 2000 cases in order to be able to perform sub-type analyses.

Statistical analysis

Relationship between case/control status and each exposure variable and 95% confidence intervals will be individually estimated using conditional logistic regression models. If necessary, a multilevel logistic model will be implemented in order to take into account the data variability due to the multicenter design and the various CRA that will administer the questionnaires (even if each CRA will be trained to the administration of the questionnaires).

As previously mentioned, previous case-control studies had limited sample sizes leading to insufficient power to detect small but relevant increases in risk. As a consequence, sarcomas were studied as a single outcome and not by histological sub-types; at most they were segregated into bone sarcomas and soft-tissue sarcomas. Such analyses supported the strong hypothesis that the same aetiology is shared between each sub-type and each site. In this study, we plan to perform stratified histological sub-types analyses. Moreover, since sarcomas may be classified into four groups on a molecular basis (i.e. sarcomas with recurrent translocation, sarcomas with specific activating or inactivating mutations, sarcomas

with MDM2 amplifications, and sarcomas with a complex genomic profile), we will explore the feasibility of classifying sarcomas by genetic types instead of by histological subtypes for the objective of identifying environmental risk factors. The underlying hypothesis is that the aetiology among these four molecular groups may be homogeneous.

Methods for coordinating the project and for quality control.

This study is organized around three different centers or committees: a national coordination center, a steering committee, and local centers.

The national coordination center will be in charge of running routine operations, assisting participating local centers, guaranteeing the quality of the data collection, centralizing data, supervising the data coding and exposure assessment. Members of this national coordination center will meet on a regular basis (i.e. once a week) to ensure reactivity to deal with emerging problems and to ensure successful project advancement. Exceptional meetings will be planned if necessary.

The steering committee's principal role will be to establish research priorities based on the availability of the data and the current scientific knowledge. The executive committee will meet annually to establish or support research project. This committee may be supplemented by external scientific members in order to obtain advice on specific questions.

Local centers will be in charge of determining identification sources of cases in their districts and of collecting data. They will also transmit the collected data to the national coordination center. Local centers will include a CRA under the directory of a coordinator.

Standardized procedures are written in a procedure manual to specify information circuit and to guarantee the quality of the collected data. Besides, a completion guide of the standardized questionnaire has been developed to assist CRA during interviews. These procedure manual and completion guide might evolve during the study period to deal with emerging problems not planned at the protocol step. A data manager will ensure that the

study is conducted in compliance with the protocol. On a regular basis and as frequently as necessary, he/she will assess the quality of the collected data using several indicators: the average time spent by the CRA at the subject's home, the degree of completeness of the questionnaires by CRA, the ratio of included cases to expected cases, the average elapsed time between the identification of a case and the interview, the average elapsed time between case and control interviews, etc. Besides, the data manager will detect aberrant data, duplication, inaccurate and missing data. Every week, this data manager will refer to the head of the national coordination center and report on the progress of the study.

All collected data will be centralized in a single location, the University of Bordeaux. These data will be informatized by a contractor specialized in health and exposure data coding (CREDIM, Centre de Recherche et de Développement en Informatique Médicale).

Expected results, as well as possible spin-offs for them.

This research is an innovative study that will permit to improve scientific knowledge about the aetiology of sarcomas and we expect our results to contribute to better understand the causes of sarcomas. We will primarily address the question of environmental causes, which will permit to confirm or generate new hypotheses on the environmental risk factors of this disease. This study will also provide a unique platform with data that will permit to address several research questions other than environmental ones.

This project will also contribute to increase the French expertise with regards to research and management of sarcomas and will reinforce existing collaboration at a national level between each team involved in this study. We expect that this study will be a starter for other studies at the european / international level in order to create an international consortium which will pool original individual-level data with harmonized data collection and exposure information.

Patient and Public Involvement

The development of the design of the study was developed without patients or public. However, before subject's inclusion, patient associations will be informed about setting up of the study and an information note will be post in the prefecture of the participating district

ETHICS AND DISSEMINATION

The present study is promoted by the French National Institute of health and medical research

The present protocol has been approved by an independent French Ethic Committee under the identification number 2018-A00519-46 and by the National Data Protection Commission under the identification number 918171.

Research outputs from this study will be disseminated through presentations at national and international conferences or workshops and through scientific publications in peer-reviewed journals.

AUTHOR CONTRIBUTIONS

All authors were involved in the development and the writing of the protocol.

FUNDING STATEMENT

This work was supported by the French National Cancer Institute in the framework of the SHS-E-SP grant - grant number 2016-118 and by the French National Institute of health and medical research in the framework of the environment and cancer grant – grant number 17CE037-00.

COMPETING INTERESTS STATEMENT

The authors declare no conflict of interest.

REFERENCES

1. Burningham Z, Hashibe M, Spector L, Schiffman JD. The epidemiology of sarcoma. Clinical sarcoma research 2012;**2**(1):14 doi: 10.1186/2045-3329-2-14[published Online First: Epub Date]|.
2. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2012, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2012/, based on November 2014 SEER data submission, posted to the SEER web site, April 2015. 2015
3. Guillou L, Aurias A. Soft tissue sarcomas with complex genomic profiles. Virchows Archiv : an international journal of pathology 2010;**456**(2):201-17 doi: 10.1007/s00428-009-0853-4[published Online First: Epub Date]|.
4. Lurkin A, Ducimetiere F, Vince DR, et al. Epidemiological evaluation of concordance between initial diagnosis and central pathology review in a comprehensive and prospective series of sarcoma patients in the Rhone-Alpes region. BMC Cancer 2010;**10**:150 doi: 10.1186/1471-2407-10-150[published Online First: Epub Date]|.
5. Ducimetiere F, Lurkin A, Ranchere-Vince D, et al. Incidence of sarcoma histotypes and molecular subtypes in a prospective epidemiological study with central pathology review and molecular testing. PloS one 2011;**6**(8):e20294 doi: 10.1371/journal.pone.0020294[published Online First: Epub Date]|.
6. Mathoulin-Pelissier S, Chevreau C, Bellera C, et al. Adherence to consensus-based diagnosis and treatment guidelines in adult soft-tissue sarcoma patients: a French prospective population-based study. Ann Oncol 2014;**25**(1):225-31 doi: 10.1093/annonc/mdt407[published Online First: Epub Date]|.
7. Wibmer C, Leithner A, Zielonke N, Sperl M, Windhager R. Increasing incidence rates of soft tissue sarcomas? A population-based epidemiologic study and literature review. Ann Oncol 2010;**21**(5):1106-11 doi: 10.1093/annonc/mdp415[published Online First: Epub Date]|.
8. Howe HL, Wingo PA, Thun MJ, et al. Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst 2001;**93**(11):824-42
9. Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978-2001: An analysis of 26,758 cases. Int J Cancer 2006;**119**(12):2922-30 doi: 10.1002/ijc.22239[published Online First: Epub Date]|.
10. Ray-Coquard I, Montesco MC, Coindre JM, et al. Sarcoma: concordance between initial diagnosis and centralized expert review in a population-based study within three European regions. Ann Oncol 2012;**23**(9):2442-9 doi: 10.1093/annonc/mdr610[published Online First: Epub Date]|.
11. Karanian M, Coindre JM. [Fourth edition of WHO classification tumours of soft tissue]. Annales de pathologie 2015;**35**(1):71-85 doi: 10.1016/j.annpat.2014.11.003[published Online First: Epub Date]|.
12. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100. A review of human carcinogens, part D: radiation. Lyon: IARC (International Agency for Research on Cancer), 2012.
13. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100. A review of human carcinogens, part F: chemical agents and related occupations. Lyon: IARC (International Agency for Research on Cancer), 2012.
14. Blair A, Zahm SH. Agricultural exposures and cancer. Environmental health perspectives 1995;**103 Suppl 8**:205-8
15. Eriksson M, Hardell L, Adami HO. Exposure to dioxins as a risk factor for soft tissue sarcoma: a population-based case-control study. Journal of the National Cancer Institute 1990;**82**(6):486-90

16. Hardell L, Eriksson M. The association between soft tissue sarcomas and exposure to phenoxyacetic acids. A new case-referent study. Cancer 1988;**62**(3):652-6
17. Hoppin JA, Tolbert PE, Flanders WD, et al. Occupational risk factors for sarcoma subtypes. Epidemiology 1999;**10**(3):300-6
18. Kogevinas M, Kauppinen T, Winkelmann R, et al. Soft tissue sarcoma and non-Hodgkin's lymphoma in workers exposed to phenoxy herbicides, chlorophenols, and dioxins: two nested case-control studies. Epidemiology 1995;**6**(4):396-402
19. Smith JG, Christophers AJ. Phenoxy herbicides and chlorophenols: a case control study on soft tissue sarcoma and malignant lymphoma. Br J Cancer 1992;**65**(3):442-8
20. Woods JS, Polissar L, Severson RK, Heuser LS, Kulander BG. Soft tissue sarcoma and non-Hodgkin's lymphoma in relation to phenoxyherbicide and chlorinated phenol exposure in western Washington. Journal of the National Cancer Institute 1987;**78**(5):899-910
21. Sathiakumar N, Delzell E, Austin H, Cole P. A follow-up study of agricultural chemical production workers. American journal of industrial medicine 1992;**21**(3):321-30
22. Grimalt JO, Sunyer J, Moreno V, et al. Risk excess of soft-tissue sarcoma and thyroid cancer in a community exposed to airborne organochlorinated compound mixtures with a high hexachlorobenzene content. International journal of cancer 1994;**56**(2):200-3
23. Pahwa P, McDuffie HH, Dosman JA, et al. Exposure to animals and selected risk factors among Canadian farm residents with Hodgkin's disease, multiple myeloma, or soft tissue sarcoma. Journal of occupational and environmental medicine 2003;**45**(8):857-68 doi: 10.1097/01.jom.0000083033.56116.c1[published Online First: Epub Date]|.
24. Hoar SK, Blair A, Holmes FF, et al. Agricultural herbicide use and risk of lymphoma and soft-tissue sarcoma. Jama 1986;**256**(9):1141-7
25. Hoar Zahm S, Blair A, Holmes FF, Boysen CD, Robel RJ. A case-referent study of soft-tissue sarcoma and Hodgkin's disease. Farming and insecticide use. Scandinavian journal of work, environment & health 1988;**14**(4):224-30
26. Merletti F, Richiardi L, Bertoni F, et al. Occupational factors and risk of adult bone sarcomas: a multicentric case-control study in Europe. International journal of cancer 2006;**118**(3):721-7 doi: 10.1002/ijc.21388[published Online First: Epub Date]|.
27. Hossain A, McDuffie HH, Bickis MG, Pahwa P, Cross Canada Study of P, Health R. Case-control study on occupational risk factors for soft-tissue sarcoma. Journal of occupational and environmental medicine 2007;**49**(12):1386-93 doi: 10.1097/JOM.0b013e318157d2f2[published Online First: Epub Date]|.
28. Briggs NC, Levine RS, Hall HI, Cosby O, Brann EA, Hennekens CH. Occupational risk factors for selected cancers among African American and White men in the United States. American journal of public health 2003;**93**(10):1748-52
29. Wingren G, Fredrikson M, Brage HN, Nordenskjold B, Axelson O. Soft tissue sarcoma and occupational exposures. Cancer 1990;**66**(4):806-11
30. Landrigan PJ. Critical assessment of epidemiologic studies on the human carcinogenicity of 1,3-butadiene. Environmental health perspectives 1990;**86**:143-7
31. Ballinger ML, Goode DL, Ray-Coquard I, et al. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. Lancet Oncol 2016;**17**(9):1261-71 doi: 10.1016/S1470-2045(16)30147-4[published Online First: Epub Date]|.
32. Thomas D. Gene--environment-wide association studies: emerging approaches. Nature reviews. Genetics 2010;**11**(4):259-72 doi: 10.1038/nrg2764[published Online First: Epub Date]|.
33. Collins FS, Varmus H. A new initiative on precision medicine. The New England journal of medicine 2015;**372**(9):793-5 doi: 10.1056/NEJMp1500523[published Online First: Epub Date]|.
34. Perrier L, Rascle P, Morelle M, et al. The cost-saving effect of centralized histological reviews with soft tissue and visceral sarcomas, GIST, and desmoid tumors: The experiences of the pathologists of the French Sarcoma Group. PloS one 2018;**13**(4):e0193330 doi: 10.1371/journal.pone.0193330[published Online First: Epub Date]|.

35. Blay JY, Soibinet P, Penel N, et al. NETSARC/RREPS and French Sarcoma Group–Groupe d'Etude des Tumeurs Osseuses (GSF-GETO) networks. Improved survival using specialized multidisciplinary board in sarcoma patients. Ann Oncol 2017;**28**(11):2852-9
36. Fletcher CD, Hogendoorn P, Mertens F, Bridge J. *WHO Classification of Tumours of Soft Tissue and Bone. 4th ed* Lyon, France: IARC Press, 2013.

BMJ Open

The Etiosarc Study : environmental aetiology of sarcomas from a French prospective multicentric population-based case-control study - study protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-030013.R1
Article Type:	Protocol
Date Submitted by the Author:	12-Apr-2019
Complete List of Authors:	Lacourt, Aude; Université de Bordeaux, Inserm U1219 - EPICENE team Amadéo, Brice; Université de Bordeaux, Inserm U1219 - EPICENE team; Université de Bordeaux, Inserm CIC 1401, Registre des cancers de la Gironde Gramond, Céline; Université de Bordeaux, Inserm U1219 - EPICENE team Marrer, Emilie; Groupe Hospitalier de la Région de Mulhouse et Sud Alsace, Registre des cancers du Haut-Rhin Plouvier, Sandrine; Registre des cancers de Lille et de sa Région Baldi, Isabelle; Université de Bordeaux, Inserm U1219 - EPICENE team; CHU de Bordeaux, Pôle de Santé Publique, Service de Médecine du Travail et Pathologies professionnelles Blay, Jean-Yves; Université Claude Bernard Lyon 1, Department of Medical Oncology, Centre Léon Berard Coindre, Jean-Michel; Institute Bergonie, Department of Biopathology de Pinieux, Gonzague; CHRU Tours, service d'anatomopathologie Gouin, François; CHU Nantes, Inserm UI957 Italiano, Antoine; Institute Bergonie, Department of Medical Oncology Le Cesne, Axel; Institut Gustave Roussy, Medical Oncology Department Le Loarer, François; Institut Bergonié, Department of Biopathology Monnereau , Alain; Université de Bordeaux, Inserm U1219 - EPICENE team; Institute Bergonie, Gironde registry of haematological malignancies Pellegrin, Isabelle; CHU de Bordeaux, Service d'Immunologie-Immunogénétique, Pôle de Biologie et Pathologie Penel, Nicolas; Ctr Oscar Lambret, Medical Oncology Department Ray-Coquard, Isabelle; Université Claude Bernard Lyon 1, Department of Medical Oncology, Centre Léon Berard Toulmonde, Maud; Institute Bergonie, Department of Medical Oncology Ducimetière, Françoise ; Centre Léon Bérard, Réseau de référence sarcome Mathoulin-Pélissier , S; Université de Bordeaux, Inserm U1219 - EPICENE team; Institute Bergonie, Clinical and Epidemiological Research Unit, Inserm CIC 1401
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Occupational and environmental medicine

Keywords:	cancer, Sarcoma < ONCOLOGY, case-control study, environmental and occupational exposure, aetiology

The Etiosarc Study: environmental aetiology of sarcomas from a French prospective multicentric population-based case-control study - study protocol

Lacourt A[1], Amadéo B[1,2], Gramond C[1], Marrer E [3], Plouvier S[4], Baldi I[1,5], Blay JY[6], Coindre JM[7], De Pinieux G[8], Gouin F[9], Italiano A[10],Le Cesne A[11], Le Loarer F[7], Monnereau A[1,12], Pellegrin I[13], Penel N[14], Ray-Coquard I[6], Toulmonde M[10], Ducimetière F[15], Mathoulin-Pélissier S[1,16] for the Etiosarc study group

1 Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, Epicene team, UMR 1219, F-33000 Bordeaux, France

2 Univ Bordeaux, Inserm CIC1401, Registre des cancers de la Gironde, Bordeaux, France

3 Registre des cancers du Haut-Rhin, Groupe hospitalier de Mulhouse et Sud Alsace (GHRMSA), Mulhouse, France

4 Registre Général des Cancers de Lille et de sa Région, Lille, France

5 CHU de Bordeaux, Pôle de Santé Publique, Service de Médecine du Travail et Pathologies professionnelles, Bordeaux, France

6 Université Claude Bernard Lyon I, Department of Medical Oncology, Centre Leon Berard, Lyon, France

7 Department of Biopathology, Institute Bergonié, Bordeaux, France

8 Department of Pathology, CHU de Tours, Tours, France

9 CHU Hôtel-Dieu, Nantes, France; INSERM UI957, Nantes, France

10 Department of Medical Oncology, Institute Bergonié, Bordeaux, France

11 Department of Medical Oncology & Surgery, Institut Gustave Roussy, Villejuif, France

12 Gironde registry of haematological malignancies, Institut Bergonié, F-33000 Bordeaux, France
13 Service d'Immunologie-Immunogénétique, Pôle de Biologie et Pathologie, Groupe Hospitalier Pellegrin, CHU Bordeaux, Bordeaux, France
14 Medical Oncology Department, Centre Oscar Lambret, Lille, France
15 Réseau de référence sarcome, Centre Léon Bérard, Lyon, France
16 Clinical and Epidemiological Research Unit, INSERM CIC1401, Institut Bergonié, Comprehensive Cancer Center, F-33000 Bordeaux, France

Etiosarc study group : Lacourt A, Amadéo B, Gramond C, Marrer E, Plouvier S, Baldi I, Bara S, Bazille C, Blay JY, Bompas E, Chaigneau L, Chateau MC, Coindre JM, Coureau G, Cupissol D, D'Almeida T, Defossez G, Delafosse P, Delcambre Lair C, De Pinieux G, Di Marco A, Fabre T, Fiorenza F, Ghnassia JP, Gouin F, Guizard AV, Italiano A, Kurtz JE, Lebrun-Ly, Le Cesne A, Le Loarer F, Le Nail LR, Maynou C, Missenard G, Molinié F, Monnereau A, Moreau A, Penel N, Ranchère-Vince D, Ray-Coquard I, Robin YM, Terrier P, Toulmonde M, Tretarre B, Velten M, Woronoff AS, Ducimetière F, Mathoulin-Pélissier S.

CORRESPONDING AUTHOR

Aude Lacourt
Equipe EPICENE Epidemiology of cancer and environmental exposures
Bordeaux Population Health Center - Inserm U1219
Université Bordeaux - ISPED
146 rue Leo Saignat
33076 Bordeaux Cedex - France
Tel:+33 5 57 57 16 23 - Fax: +33 5 57 57 47 33
aude.lacourt@inserm.fr

ABSTRACT

Introduction: Sarcomas are rare tumors of connective tissue. The exact overall incidence of sarcomas is unknown due to diagnostic difficulties and the various histological subtypes (over 80 subtypes). However, the apparent increasing incidence of sarcomas suggests environmental causes such as pesticides. Except for some specific factors (i.e. ionizing radiation, vinyl chloride, dioxin, and genetic predispositions) the scientific knowledge on the aetiology of sarcomas is sparse and inconsistent. France is a particularly appropriate country to set up a study investigating the causes of sarcoma occurrence due to the French organization in treatment and care of sarcoma patients, which is highly structured and revolved around national expert networks. The main objective of the ETIOSARC project is to study the role of lifestyle, environmental and occupational factors in the occurrence of sarcomas among adults from a multicentric population-based case-control study.

Methods and analysis: Cases will be all incident patients (older than 18 years old) prospectively identified in 15 districts of France covered by a general population-based cancer registry and/or a reference center in sarcoma's patient care over a three-year period with an inclusion start date ranging from April 2019 to January 2020 and histologically confirmed by a second review of the diagnosis. Two controls will be individually-matched by sex, age (5-years group), and districts of residence and randomly selected from electoral rolls. A standardized questionnaire will be administered by a trained interviewer in order to gather information about occupational and residential history, demographic and socioeconomic characteristics and lifestyle factors. At the end of the interview, a saliva sample will be systematically proposed.

This study will permit to validate or identify already suspected risk factors for sarcomas such as phenoxyherbicides, chlorophenol and to generate new hypothesis to increase our understanding about the genetic and environmental contributions in the carcinogenicity process.

Ethics and dissemination: The present study is promoted by the French National Institute of Health and Medical Research (identification number C17-03).

This study received National French Ethic committee (CPP Sud Méditerrannée I) approval (identification number 18-31) and French Data Protection Authority (CNIL) approval (identification number 918171). Results of this study will be published in international peer reviewed journals.

Technical appendix, statistical code, and dataset will be available in the Dryad repository when collection data is completed.

Trial registration number: NCT03670927

Count words: 4629

Keywords: cancer, sarcoma, case-control study, environmental and occupational exposure, aetiology

Article Summary

Strengths and limitations of this study

- All newly diagnosed patients will be ascertained through a systematic review of the diagnostic by an expert pathologist within French sarcomas networks.
- Prospective inclusion of more than 2000 population-based patients which will allow to perform analysis by histological subtypes.
- Inclusion of more than 2000 population-based patients which will allow to perform analysis by histological subtypes.
- Collection of various exposure data using face-to face administered standardized questionnaire by trained interviewers. Additionally, saliva samples will be collected
- Possible low participation rate of population controls.
- Retrospective collection of data and over a long period of time.

INTRODUCTION

Sarcomas are a heterogeneous group of rare malignant tumors of connective tissues. Besides multiple and complex histology (approximatively over 80 subtypes), these tumors can occur in almost any anatomic site. It is usual to distinguish bone sarcomas (osteosarcomas and chondrosarcomas) from soft-tissue sarcomas (muscles, joints, fat, nerves, skin tissues and blood vessels) and visceral sarcomas. Even though it accounts for less than 1% of adult cancers, sarcoma is one of the most frequent cancer types in young adults [1]. Based on statistics of the Surveillance Epidemiology and End Results (SEER) program from 2008-2012, median age at diagnosis for soft-tissue sarcomas and bones sarcomas were 59 and 42 years of age, respectively and median age at death was 65 years old for soft-tissue sarcomas and 59 years old for bones sarcomas [2].

At the genetic level, sarcomas can be split into two large categories based on the rearrangement level of the genome. On the one side sarcomas with very simple genetics (50% of all sarcomas) based on point mutation (gastrointestinal stromal tumor (GIST), desmoïd tumor) or a specific and recurrent translocation (Ewing, Synovial sarcoma, Myxoïd liposarcoma, …); and on the other side, sarcomas with a very complex genetics (50% of all sarcomas, leiomyosarcoma, angiosarcoma, undifferentiated pleomorphic sarcoma,…) [3].

The incidence of sarcomas has been very imperfectly estimated due to both diagnostic confusion with carcinomas of the same organ and the variety of localization of these tumors. Thirty percent of sarcomas are misclassified at initial diagnosis [4]. Cancer incidence is often reported by site possibly leading to an under-estimation of incidence for some subtypes. In France, the world age-standardized incidence rates of overall sarcomas was estimated at 4.8 per 100,000 inhabitants per year [5] and soft-tissue sarcoma's incidence, which account for more than 50% of sarcomas patients was estimated at 3.3 per 100,000 inhabitants per year [6]. While some authors did not confirm an incidence increase over last years [7], others reported a statistically significant raise of the incidence of sarcomas [8 9] leading to the

hypotheses of an implication of environmental factors in the aetiology of this cancer. However, the role of diagnostic and reporting practice cannot be ruled out.

The various histological subtypes, various anatomical sites as well as the rarity of the disease make the aetiology of this cancer difficult to study [1]. Some studies have attempted to investigate the role of some environmental factors; however, results are often inconclusive or inconsistent from one study to another. Thus, to date, it is impossible to draw strong conclusions on the aetiology of sarcomas from existing studies. Indeed, there are some methodological considerations to explain discrepancies between studies: sample sizes are often limited leading to insufficient power to detect small but relevant increases in risk. As a consequence, sarcomas are studied as a single outcome and not by histological sub-types; at most they are segregated into bone sarcomas and soft-tissue sarcomas. Such analyses support the strong hypothesis that the same aetiology is shared between each sub-type and each site. The inclusion periods of patients are usually old and ascertainment of patients may be incorrect. Indeed, a second expert review of diagnosis is essential to correctly classify sarcoma's tumors [10]. Diagnostic procedures have been refined in expert centers during the last several years with the inclusion of new genetic and molecular data leading to a better understanding and definition of tumors. As a consequence, a new World Health Organization classification of soft-tissue tumors was published in 2013 [11].

From IARC evaluation, the strongest evidence for environmental cause is for ionizing radiations (including radiotherapy) for both soft tissue and bone sarcomas that are classified in group 1 as carcinogenic to humans [12]. Convincing evidence also exists for linking vinyl chloride to the occurrence of a specific type of sarcoma (liver angiosarcoma). This conclusion arises from consistent observations in vinyl chloride industry that conducted IARC to classify this agent also in group 1 as carcinogenic in humans [13]. The dioxin 2,3,7,8-Tetrachlorodibenzo-p-dioxin (2,3,7,8-TCDD), released to the environment during the combustion of fossil fuels and wood, and waste incineration has also been classified carcinogenic by the IARC, with evidence for soft-tissue sarcoma [13].

Because epidemiological studies have consistently demonstrated higher mortality and incidence rates from soft-tissue sarcomas in farmers, specific attention has been paid to the potential role of pesticide exposure in the occurrence of these tumors [14]. Several case-control studies assessed the relationship between herbicide and chlorophenol with soft-tissue sarcoma risk but results varied from no association to a strong relationship for exposed populations [15-20]. Discrepancies between studies may be due in part to the low statistical power due to small sample sizes but most importantly to difficulties in exposure assessment. Indeed, disentangling the various contributions of chlorophenols, herbicides and dioxins is complicated because chlorophenols are used in the production of herbicides and dioxins are also a contaminant in herbicides production.

Some studies have also suggested that other pesticides than phenoxys could play a role in the occurrence of sarcomas. Increased risk of soft-tissue sarcoma have been observed in a cohort of workers in an organochlorine production plant [21], in a population living in the vicinity of an organochlorinated-compounds factory in Spain [22], and also in a case-control study in Canada exploring some specific pesticides [23]. This last study found significant associations with two insecticides: Aldrin (OR=3.71, 1.00-13.76) and Diazinon (OR=3.31, 1.78-6.23) and a trend with formaldehyde (OR=2.07, 0.94-4.56). A case-control study in Kansas failed to find associations with pesticides used on crops, including herbicides [24] but found an increased risk with the use of insecticides to animals [25], higher for farmers who mixed or applied themselves and for those who did not use any protective equipment, also more pronounced for so-called “fibrous” and “myomatous” sarcomas. A European study on risk of adult bone sarcomas also found an association with pesticides that was similar for insecticides and herbicides [26].

Other occupational exposures have also been explored and, besides farming, some industries have been associated in some studies with elevated risks of soft-tissue sarcomas: gardeners, meat packers, sawmill workers, machinists, ground maintenance workers [27-29]. Exploring specific chemicals or agents, positive associations have been found with wood

dust [28], radium [27] and 1-3 butadiene [30] for soft-tissue sarcoma. A multicenter case-control study in seven European countries specifically focused on bone sarcomas found increased risks among blacksmiths, toolmakers, machine-tool operators and construction workers [26].

Besides occupational factors, few studies investigated other environmental factors related to lifestyle, including, the potential role of female hormones, tobacco smoke, alcohol, nutrition, fluoride in drinking water, and body mass index. These studies reported inconclusive or inconsistent results with the exception of viruses like the human immunodeficiency virus (HIV) and human herpes virus 8 (HHV8) for Kaposi's sarcoma [1].

Finally, while there is increased evidence regarding genetic determinants of sarcoma risk [31], and growing evidence that gene × environment (G × E) interactions are determinants of development and progression of complex disease [32], to date, there is no data related to such G × E interactions with regards to sarcoma risk. However, studying G × E interactions may help to identify susceptible groups of individuals which is essential to better target prevention programs or to develop precision medicine [33].

Several hypotheses have been studied regarding the aetiology of sarcomas. Nevertheless, the methodological limitations such as diagnosis certification, small sample sizes, exposure assessment methodology, and the analysis strategies that do not distinguish the various histological subtypes making the hypothesis that the aetiology is homogeneous across all histological subtypes prevent any definite conclusions. There is a clearly identified need to further investigate the aetiology of sarcomas with improved study designs including a systematic centralized diagnosis review by a second expert, increased sample sizes, more refined exposure assessment methods and also a biological component to increase our understanding of biologic mechanisms of carcinogenesis and to study the interaction between the genetic and environment component.

France is a particularly appropriate country to set up such study due to the French organization in treatment and care of sarcoma patients around three national networks

labelled by the French National Cancer Institute (INCa): the network of reference for soft-tissue sarcoma pathology (RRePS: https://rreps.sarcomabcb.org/home.htm), the network of reference for bone sarcoma pathology (ResOs: https://resos.sarcomabcb.org/home.htm) and the clinical sarcoma network (NetSarc: https://netsarc.sarcomabcb.org/home.htm). The objectives of RRePS and ResOs are to ensure systematic and free secondary reviews for all new diagnoses of soft tissue and visceral sarcomas, gastro-intestinal stromal tumors (GIST), desmoid tumors and bone sarcomas across the whole of France, and to facilitate access to molecular biology analyses, collection of samples for biological resource centers, to participate in research and clinical trials, to draft good clinical practice guidelines for professionals and information documents for patients, and to organize continuing education and information for patients [34]. The NetSarc clinical network for sarcoma is the clinical network of the French Sarcoma Group dedicated to clinical patient care. NetSarc is managed by three sites (Centre Leon Berard in Lyon, Institut Bergonie in Bordeaux and Institut Gustave Roussy in Villejuif) working closely with 25 expert regional centers, ensuring good coverage of the whole French territory [35]. Besides, the French population-based cancer registries are organized in a collaborative network named Francim. The main objectives of this network are to coordinate the 14 general cancer registries and 11 specialized cancer registries that exhaustively register all newly diagnosed and confirmed cancer patients according to international procedures, to harmonize patients registration and data quality, to provide epidemiological indicators (incidence, survival, prevalence) and coordinate epidemiological and surveillance research on cancer.

The main objective of the ETIOSARC study is to assess the role of lifestyle, environmental and occupational factors in the occurrence of sarcomas among adults from a multicenter population-based case-control study.

Specific objectives are:

- To identify environmental risk factors for sarcomas as a whole and for the most frequent subtypes;

- To investigate the interactions between gene polymorphisms and environmental exposures in sarcoma susceptibility;
- To assess whether some specific genetic characteristics of sarcoma' tumors are associated with environmental exposures;

We will also explore the feasibility of classifying sarcomas by genetics types (simple vs complex genomic profile) instead of by histological subtypes as part of the objective of identifying environmental risk factors.

METHODS AND ANALYSIS

Study design

The ETIOSARC study is a prospective multicenter population-based case-control study. This study is restricted to French geographical areas (further called districts) that meet at least one of these four criteria (Figure 1):

- Criteria 1: Districts covered by both a general cancer registry and a French Sarcoma Group (GSF-GETO) expert center from the sarcoma reference network;
- Criteria 2: Districts including a coordinator center of any of the RRePS / NetSarc networks;
- Criteria 3: Districts covered by a general cancer registry that is expected to register more than 50 patients per year;
- Criteria 4: Districts adjacent to districts meeting criteria 1 and covered by a general cancer registry.

In total, 15 districts meet one of the four criteria, which represent 17,813,937 inhabitants, approximatively 27% of the French population (2019 estimations from the national institute of statistics and economic studies, Insee). Eight districts meet criteria 1 (Gironde, Hérault, Haute-Vienne, Loire Atlantique, Calvados, metropolitan area of Lille, Bas-Rhin and Doubs), two districts meet criteria 2 (Rhône and Val-de-Marne); three districts meet criteria 3 (Isère, Haut-Rhin and Poitou-Charentes) and two districts meet criteria 4 (Manche, Vendée).

Since sarcomas are rare tumors, this study has to be multicentric. Indeed, from general cancer registry data and RRePS overall incidence data, considering a patients' response rate of 70%, and considering that approximatively 90% of newly diagnosed patients will benefit from a systematic secondary review of diagnosis, it is expected to include 718 incident patients per year from these 15 districts, corresponding to a total sample size of 2 154 patients from a 3 years-recruitment (Table1). The study is planned to start in April 2019 and to end in April 2022.

< Table 1: Total number of expected patients includes in the Etiosarc study over a three years recruitment >

	Expected number of new sarcoma patients / year *	Expected number of confirmed patients / year†	Expected number of interviewed patients /year‡	Expected total number of included patients¶
Criteria 1: Districts covered by both a general cancer registry and a French Sarcoma Group (GSF-GETO) expert center from the sarcoma reference network				
Gironde	120	108	76	228
Hérault	86	77	54	162
Haute-Vienne	23	21	15	45
Calvados	50	45	31	93
Loire-Atlantique	98	88	62	186
Metropolitan area of Lille	48	43	30	90
Bas-Rhin	90	81	57	171
Doubs	42	38	27	81
Criteria 2: Districts including a coordinator center of any of the RRePS / NetSarc networks				
Rhône	116	104	73	219
Val de Marne	68	61	43	129
Criteria 3: Districts covered by a general cancer registry that is expected to register more than 50 patients per year				
Isère	89	80	56	168
Haut-Rhin	56	50	35	105
Poitou-Charentes	167	150	105	315
Criteria 4: Districts adjacent to districts meeting criteria 1 and covered by a general cancer registry				
Vendée	51	46	32	96
Manche	34	31	22	66
Total	**1 138**	**1 023**	**718**	**2 154**

* Estimation from general cancer registry data (except for Rhône et Val de Marne, estimation from RRePS data).

† Estimation with a systematic secondary review a diagnostic confirmation for 90% of patients.

‡ Estimation with a response rate of 70%.

¶ Over a three years study.

Study population

Cases definition and recruitment modalities

Cases are defined as all incident patients with a diagnosis of primary sarcoma and histologically confirmed by an expert pathologist of the RRePS or ResOs networks in the 15 districts of France participating to this study.

Inclusion criteria are:

- Patients diagnosed in the previous 6 months from identification with a primary and histologically confirmed malignant sarcoma including soft-tissue, visceral and bone sarcomas as defined by the WHO classification of bones and soft tissue sarcoma, fourth edition, 2013 [36];
- Diagnosed over a three-year period with an inclusion start date ranging from the 1st February 2019 to the 1st January 2020 depending on the districts;
- Living in one of the 15 districts participating to the study at the time of diagnosis;
- At least 18 years old at diagnosis;
- Agreed to participate to the study with a signed informed consent;

Non-inclusion criteria are:

- Patients with a known genetic predisposition to sarcoma such as Li-Fraumeni syndrome, Retinoblastoma syndrome, neurofibromatosis;
- Kaposi's' sarcoma;
- Protected adults' patients (aged of at least 18 years old) under guardianship by court order.

Patients will be recruited by specifically trained clinical research associate (CRA). In districts where a general cancer registry is active, the process of identifying incident sarcoma patients is well defined, which warrants the efficiency and exhaustiveness of the recruitment. However, poor survival for some cancer patients will not allow to rely on routine inclusion procedure and the registries will implement a rapid patient ascertainment procedure to minimize the delay between diagnosis and enrollment and interview. Patients will be identified from pathology laboratories and multidisciplinary sarcoma tumor board (including sarcoma, gynecological, digestive, skin and bone tumor board). CRA will regularly contact laboratories and hospitals (within a three months window) to identify newly diagnosed sarcoma patients and to collect associated reports. The CRA will retrieve the clinical file in order to check inclusion and non inclusion criteria and to collect names and addresses of eligible patients and their physicians. Then, the CRA will contact the patients' physicians in order to gather their clinical advice to include their patients in the study. In case of agreement, the CRA will contact the patients and ask them to participate. In case of oral agreement, he/she will arrange an interview to collect written consent and administer detailed questionnaire.

In Rhône and Val de Marne where no general cancer registry exists, the process of identifying incident sarcoma patients will largely rely on the existence of a sarcoma reference network coordination center to facilitate the identification of new patients.

Regardless of the district, the diagnosis of all included patients will be systematically ascertained by an expert pathologist within the RRePS and ResOs networks following a standard procedure.

On a regular basis (at least once a month), the list of incident patients included in the RRePS/ResOS/NetSarc networks will be extracted from the shared databases. The list of patients identified by registries will be merged with the list of patients included in the RRePS/ResOS/NetSarc networks by the pathology sample reference number in order to collect diagnosis confirmation and to identify new but missed incident patients.

Controls selection

Two control subjects per patient will be randomly selected from the French general population using electoral rolls and individually matched to patient by age (by 5-year age group), sex and residential area (French department).

Inclusion criteria are:

- Subjects registered within the electoral rolls;
- At least 18 years old at interview;
- Living in one of the 15 districts participating to the study at the time of interview;
- Agreed to participate to the study with a signed informed consent;

Non-inclusion criteria are:

- Subject previously diagnosed with a primary and histologically confirmed malignant sarcoma including soft-tissue, visceral and bone sarcomas as defined by the WHO classification of bones and soft tissue sarcoma, fourth edition, 2013;
- Protected adults' patients (aged of at least 18 years old) under guardianship by court order.

Controls will be recruited according to an incidence density sampling procedure which will guarantee the recruitment of controls from the same source population as patients. The selection of controls and the recruitment of patients will take place simultaneously. Every time a patient will be diagnosed and identified, two controls will be randomly selected at that time from the electoral rolls of the same district as the patient district of residence. For each patient, a list of 20 potential controls with the same age (within a 5-year age group) and sex as the patients will be constituted and randomly selected in order to account for potential refusal. Calls to contact potential controls will be centralized and will be made by an investigator specifically trained to better understand the objectives of the study. Potential

controls from the constituted list will be contacted one after the other until identifying two controls per case agreeing to participate to the study. First, this investigator will try to retrieve the phone number of the first potential control on the list and will contact him/her in order to gather his/her agreement to participate in this study. After five calls made at different schedules (e.g. in the evening on weekdays), if the potential control couldn't be reached, the investigator will send an information letter containing a reply coupon indicating phone number and schedule at which the person could be reach. If the reply coupon is not returned, or after 10 calls made at the phone number and the schedule indicated on the reply coupon, the first potential control will be considered as unreachable and then, the same procedure will be applied for the second potential control on the list and so on until two controls per patient agreed to participate to the study.

For each control agreeing to participate, an appointment will be made to collect written consent and to administer detailed questionnaire by a CRA.

All participants will give their informed written consent to participation, in line with French ethical guidelines.

Data collection

Data will be collected in a three-step procedure and from subjects themselves: i) the interviewer will contact subjects by phone and ask them to participate in the study; ii) if the subject has agreed to participate, an appointment will be made and a consent form as well as a self-administered questionnaire will be sent by mail to each subject. This self-questionnaire will permit to gather the complete occupational history (for each job held for at least 6 months) and residential history (for each place occupied for at least 1 year); iii) during a face-to-face interview, the trained interviewer will check (complete if necessary) and supplement the self-administered questionnaire by a specific questionnaire with questions about demographic and socioeconomic characteristics, occupations of spouse and parents,

leisure-time activities, reproduction, medical history, family history of cancer, diet, lifestyle factors such as tobacco smoking and alcohol consumption. The specific questionnaire will also collect additional occupational and residential information such as work tasks, work places, materials handled for each job held for at least 6 months and description of the environment of each residence places.

At the end of the interview, subjects (both patients and controls) will be invited to provide a saliva sample in order to obtain germline DNA.

In case of refusal to participate and if the subject agreed, data on the last occupied job and educational level will be collected in order to assess the potential selection bias that might occur due to the specific profile of non-respondent subjects.

To ensure consistency in data collection, all CRA will be trained in the completion of the questionnaire and will detain a field guide of completion. Additionally, phone meetings involving all CRA will take place each month in order to deal with recurring problems in the completion of the questionnaire and to ensure homogeneity in data collection between all CRA. Moreover, these meetings will help maintain CRAs' motivation and level of training.

The completeness of questionnaires and the quality of interviews will be routinely checked. All questionnaires will be scanned in order to facilitate storage and possible return to the questionnaires to allow quality checks.

Biological sampling and storage

Each participant will be asked to provide a salivary sample in order to obtain germline DNA. Salivary samples will be collected by the subjects themselves under instructions from the CRA, using Genefix™ saliva DNA collection kit. After collection, samples will be sent at the Biological Resources Center of the Bordeaux hospital university center "Bordeaux Biothèque Santé" for storage (NFS-96900 certification, BBMRI-ERIC ID: FR_BB-0033-00094).

Determination of the sample size

The main objective of this study is to examine the association between environmental exposures (including general environment, occupational environment and lifestyle) and risk of sarcoma occurrence. Since our definition of environmental exposures is very broad, we have based our sample size calculation on various scenarios of exposure prevalence, from 5% (relevant for domestic, environmental but also some occupational exposures such as farmer in the general population) to 20% (relevant for some occupational exposures such as fibers). For a total sample size of 2 000 patients and a 1:2 individually-matched design, considering a statistical power of 80% at a significance level of 5%, the minimum detectable odds ratio will be 1.39 and 1.21 under an exposure prevalence of 5% and 20%, respectively. Considering sub-types analyses, the four main histological types are GIST, liposarcoma, leiomyosarcoma and unclassified sarcoma, which account for 18%, 15%, 11% and 16% of sarcomas, respectively. Considering a sample size of 300 patients, the minimum detectable odds ratio will be 2.21 and 1.61 under an exposure prevalence of 5% and 20%, respectively. Typically in environmental epidemiology, the relative increases in disease risks due to environmental exposures are usually low around 1.5, thus, it is essential to recruit a minimum of 2 000 patients in order to be able to perform sub-type analyses.

Statistical analysis

Relationship between patient/control status and each exposure variable and 95% confidence intervals will be individually estimated using conditional logistic regression models. If necessary, a multilevel logistic model will be implemented in order to take into account the data variability due to the multicenter design and the various CRA that will administer the questionnaires (even if each CRA will be trained to the administration of the questionnaires).

As previously mentioned, previous case-control studies had limited sample sizes leading to insufficient power to detect small but relevant increases in risk. As a consequence, sarcomas

were studied as a single outcome and not by histological sub-types; at most they were segregated into bone sarcomas and soft-tissue sarcomas. Such analyses supported the strong hypothesis that the same aetiology is shared between each sub-type and each site. In this study, we plan to perform stratified histological sub-types analyses. Moreover, since sarcomas may be classified into four groups on a molecular basis (i.e. sarcomas with recurrent translocation, sarcomas with specific activating or inactivating mutations, sarcomas with MDM2 amplifications, and sarcomas with a complex genomic profile), we will explore the feasibility of classifying sarcomas by genetic types instead of by histological subtypes for the objective of identifying environmental risk factors. The underlying hypothesis is that the aetiology among these four molecular groups may be homogeneous.

Sensitive analyses will be systematically implemented in order to assess the robustness of the produced results and to analyze the impact of potential bias (especially selection bias) on the produced results.

Methods for coordinating the project and for quality control.

This study is organized around three different centers or committees: a national coordination center, a steering committee, and local centers.

The national coordination center will be in charge of running routine operations, assisting participating local centers, guaranteeing the quality of the data collection, centralizing data, supervising the data coding and exposure assessment. Members of this national coordination center will meet on a regular basis (i.e. once a week) to ensure reactivity to deal with emerging problems and to ensure successful project advancement. Exceptional meetings will be planned if necessary.

The steering committee's principal role will be to establish research priorities based on the availability of the data and the current scientific knowledge. The executive committee will

meet annually to establish or support research project. This committee may be supplemented by external scientific members in order to obtain advice on specific questions.

Local centers will be in charge of determining identification sources of patients in their districts and of collecting data. They will also transmit the collected data to the national coordination center. Local centers will include a CRA under the directory of a coordinator.

Standardized procedures are written in a procedure manual to specify information circuit and to guarantee the quality of the collected data. Besides, a completion guide of the standardized questionnaire has been developed to assist CRA during interviews. These procedure manual and completion guide might evolve during the study period to deal with emerging problems not planned at the protocol step. A data manager will ensure that the study is conducted in compliance with the protocol. On a regular basis and as frequently as necessary, he/she will assess the quality of the collected data using several indicators: the average time spent by the CRA at the subject's home, the degree of completeness of the questionnaires by CRA, the ratio of included patients to expected patients the average elapsed time between the identification of a patient and the interview, the average elapsed time between patient and control interviews, etc. Besides, the data manager will detect aberrant data, duplication, inaccurate and missing data. Every week, this data manager will refer to the head of the national coordination center and report on the progress of the study.

All collected data will be centralized in a single location, the University of Bordeaux. These data will be informatized by a contractor specialized in health and exposure data coding (CREDIM, Centre de Recherche et de Développement en Informatique Médicale).

Expected results, as well as possible spin-offs for them.

This research is an innovative study that will permit to improve scientific knowledge about the aetiology of sarcomas and we expect our results to contribute to better understand the causes of sarcomas. We will primarily address the question of environmental causes, which

will permit to confirm or generate new hypotheses on the environmental risk factors of this disease. This study will also provide a unique platform with data that will permit to address several research questions other than environmental ones.

This project will also contribute to increase the French expertise with regards to research and management of sarcomas and will reinforce existing collaboration at a national level between each team involved in this study. We expect that this study will be a starter for other studies at the european / international level in order to create an international consortium which will pool original individual-level data with harmonized data collection and exposure information.

Patient and Public Involvement

The development of the design of the study was developed without patients or public. However, before subject's inclusion, patient associations will be informed about setting up of the study and an information note will be post in the prefecture of the participating district

ETHICS AND DISSEMINATION

The present study is promoted by the French National Institute of Health and Medical Research (identification number C17-03).

The present protocol has been approved by an independent French Ethic Committee (CPP Sud Méditerrannée I) under the identification number 18-31 and by the National Data Protection Commission under the identification number 918171.

Research outputs from this study will be disseminated through presentations at national and international conferences or workshops and through scientific publications in peer-reviewed journals.

Trial registration number: NCT03670927

AUTHOR CONTRIBUTIONS

AL prepared the first draft for all sections of this manuscript with the help of all co-authors.

AL, CG, BA, IB, AM, SMP participated in the literature review.

AL, BA, CG, SMP are responsible for the national coordination center and planed the first draft of the present protocol

EM, SP, BA, AM will manage the cancer registries network involved in this study

JYB, GDP, AI, ALC, NP, IRC, FD will manage the clinical center from the NetSarc clinical sarcoma network

MT is responsible of the shared databases RRePS/ResOS/NetSarc

JMC, FG, FLL are responsible of the RRePS network involved in this study

IP is responsible of the biological resources center of Bordeaux hospital university center « Bordeaux Biothèque Santé"

All authors were involved in the development and the writing of the protocol

All authors participated in the editing and correction of the final text

All authors read and approved the final manuscript.

FUNDING STATEMENT

This work was supported by the French National Cancer Institute in the framework of the SHS-E-SP grant - grant number 2016-118 and by the French National Institute of health and medical research in the framework of the environment and cancer grant – grant number 17CE037-00.

COMPETING INTERESTS STATEMENT

The authors declare no conflict of interest.

1

REFERENCES

1. Burningham Z, Hashibe M, Spector L, Schiffman JD. The epidemiology of sarcoma. Clinical sarcoma research 2012;**2**(1):14 doi: 10.1186/2045-3329-2-14[published Online First: Epub Date]|.
2. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2012, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2012/, based on November 2014 SEER data submission, posted to the SEER web site, April 2015. 2015
3. Guillou L, Aurias A. Soft tissue sarcomas with complex genomic profiles. Virchows Archiv : an international journal of pathology 2010;**456**(2):201-17 doi: 10.1007/s00428-009-0853-4[published Online First: Epub Date]|.
4. Lurkin A, Ducimetiere F, Vince DR, et al. Epidemiological evaluation of concordance between initial diagnosis and central pathology review in a comprehensive and prospective series of sarcoma patients in the Rhone-Alpes region. BMC Cancer 2010;**10**:150 doi: 10.1186/1471-2407-10-150[published Online First: Epub Date]|.
5. Ducimetiere F, Lurkin A, Ranchere-Vince D, et al. Incidence of sarcoma histotypes and molecular subtypes in a prospective epidemiological study with central pathology review and molecular testing. PloS one 2011;**6**(8):e20294 doi: 10.1371/journal.pone.0020294[published Online First: Epub Date]|.
6. Mathoulin-Pelissier S, Chevreau C, Bellera C, et al. Adherence to consensus-based diagnosis and treatment guidelines in adult soft-tissue sarcoma patients: a French prospective population-based study. Ann Oncol 2014;**25**(1):225-31 doi: 10.1093/annonc/mdt407[published Online First: Epub Date]|.
7. Wibmer C, Leithner A, Zielonke N, Sperl M, Windhager R. Increasing incidence rates of soft tissue sarcomas? A population-based epidemiologic study and literature review. Ann Oncol 2010;**21**(5):1106-11 doi: 10.1093/annonc/mdp415[published Online First: Epub Date]|.
8. Howe HL, Wingo PA, Thun MJ, et al. Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst 2001;**93**(11):824-42
9. Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978-2001: An analysis of 26,758 cases. Int J Cancer 2006;**119**(12):2922-30 doi: 10.1002/ijc.22239[published Online First: Epub Date]|.
10. Ray-Coquard I, Montesco MC, Coindre JM, et al. Sarcoma: concordance between initial diagnosis and centralized expert review in a population-based study within three European regions. Ann Oncol 2012;**23**(9):2442-9 doi: 10.1093/annonc/mdr610[published Online First: Epub Date]|.
11. Karanian M, Coindre JM. [Fourth edition of WHO classification tumours of soft tissue]. Annales de pathologie 2015;**35**(1):71-85 doi: 10.1016/j.annpat.2014.11.003[published Online First: Epub Date]|.
12. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100. A review of human carcinogens, part D: radiation. Lyon: IARC (International Agency for Research on Cancer), 2012.
13. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100. A review of human carcinogens, part F: chemical agents and related occupations. Lyon: IARC (International Agency for Research on Cancer), 2012.
14. Blair A, Zahm SH. Agricultural exposures and cancer. Environmental health perspectives 1995;**103 Suppl 8**:205-8
15. Eriksson M, Hardell L, Adami HO. Exposure to dioxins as a risk factor for soft tissue sarcoma: a population-based case-control study. Journal of the National Cancer Institute 1990;**82**(6):486-90

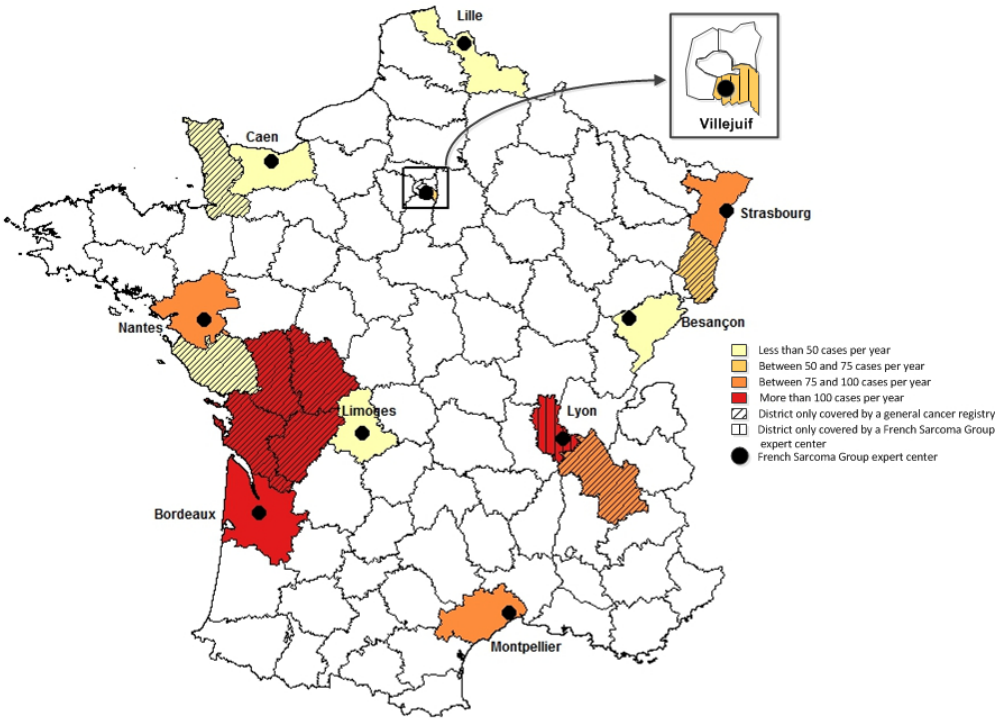
16. Hardell L, Eriksson M. The association between soft tissue sarcomas and exposure to phenoxyacetic acids. A new case-referent study. Cancer 1988;**62**(3):652-6
17. Hoppin JA, Tolbert PE, Flanders WD, et al. Occupational risk factors for sarcoma subtypes. Epidemiology 1999;**10**(3):300-6
18. Kogevinas M, Kauppinen T, Winkelmann R, et al. Soft tissue sarcoma and non-Hodgkin's lymphoma in workers exposed to phenoxy herbicides, chlorophenols, and dioxins: two nested case-control studies. Epidemiology 1995;**6**(4):396-402
19. Smith JG, Christophers AJ. Phenoxy herbicides and chlorophenols: a case control study on soft tissue sarcoma and malignant lymphoma. Br J Cancer 1992;**65**(3):442-8
20. Woods JS, Polissar L, Severson RK, Heuser LS, Kulander BG. Soft tissue sarcoma and non-Hodgkin's lymphoma in relation to phenoxyherbicide and chlorinated phenol exposure in western Washington. Journal of the National Cancer Institute 1987;**78**(5):899-910
21. Sathiakumar N, Delzell E, Austin H, Cole P. A follow-up study of agricultural chemical production workers. American journal of industrial medicine 1992;**21**(3):321-30
22. Grimalt JO, Sunyer J, Moreno V, et al. Risk excess of soft-tissue sarcoma and thyroid cancer in a community exposed to airborne organochlorinated compound mixtures with a high hexachlorobenzene content. International journal of cancer 1994;**56**(2):200-3
23. Pahwa P, McDuffie HH, Dosman JA, et al. Exposure to animals and selected risk factors among Canadian farm residents with Hodgkin's disease, multiple myeloma, or soft tissue sarcoma. Journal of occupational and environmental medicine 2003;**45**(8):857-68 doi: 10.1097/01.jom.0000083033.56116.c1[published Online First: Epub Date]|.
24. Hoar SK, Blair A, Holmes FF, et al. Agricultural herbicide use and risk of lymphoma and soft-tissue sarcoma. Jama 1986;**256**(9):1141-7
25. Hoar Zahm S, Blair A, Holmes FF, Boysen CD, Robel RJ. A case-referent study of soft-tissue sarcoma and Hodgkin's disease. Farming and insecticide use. Scandinavian journal of work, environment & health 1988;**14**(4):224-30
26. Merletti F, Richiardi L, Bertoni F, et al. Occupational factors and risk of adult bone sarcomas: a multicentric case-control study in Europe. International journal of cancer 2006;**118**(3):721-7 doi: 10.1002/ijc.21388[published Online First: Epub Date]|.
27. Hossain A, McDuffie HH, Bickis MG, Pahwa P, Cross Canada Study of P, Health R. Case-control study on occupational risk factors for soft-tissue sarcoma. Journal of occupational and environmental medicine 2007;**49**(12):1386-93 doi: 10.1097/JOM.0b013e318157d2f2[published Online First: Epub Date]|.
28. Briggs NC, Levine RS, Hall HI, Cosby O, Brann EA, Hennekens CH. Occupational risk factors for selected cancers among African American and White men in the United States. American journal of public health 2003;**93**(10):1748-52
29. Wingren G, Fredrikson M, Brage HN, Nordenskjold B, Axelson O. Soft tissue sarcoma and occupational exposures. Cancer 1990;**66**(4):806-11
30. Landrigan PJ. Critical assessment of epidemiologic studies on the human carcinogenicity of 1,3-butadiene. Environmental health perspectives 1990;**86**:143-7
31. Ballinger ML, Goode DL, Ray-Coquard I, et al. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. Lancet Oncol 2016;**17**(9):1261-71 doi: 10.1016/S1470-2045(16)30147-4[published Online First: Epub Date]|.
32. Thomas D. Gene--environment-wide association studies: emerging approaches. Nature reviews. Genetics 2010;**11**(4):259-72 doi: 10.1038/nrg2764[published Online First: Epub Date]|.
33. Collins FS, Varmus H. A new initiative on precision medicine. The New England journal of medicine 2015;**372**(9):793-5 doi: 10.1056/NEJMp1500523[published Online First: Epub Date]|.
34. Perrier L, Rascle P, Morelle M, et al. The cost-saving effect of centralized histological reviews with soft tissue and visceral sarcomas, GIST, and desmoid tumors: The experiences of the pathologists of the French Sarcoma Group. PloS one 2018;**13**(4):e0193330 doi: 10.1371/journal.pone.0193330[published Online First: Epub Date]|.

35. Blay JY, Soibinet P, Penel N, et al. NETSARC/RREPS and French Sarcoma Group–Groupe d'Etude des Tumeurs Osseuses (GSF-GETO) networks. Improved survival using specialized multidisciplinary board in sarcoma patients. Ann Oncol 2017;**28**(11):2852-9

36. Fletcher CD, Hogendoorn P, Mertens F, Bridge J. *WHO Classification of Tumours of Soft Tissue and Bone. 4th ed* Lyon, France: IARC Press, 2013.

1 FIGURE 1: districts covered by the Etiosarc study

2 Figure legend: map was created with R packages maptools, maps, raster and mapdata.



235x170mm (96 x 96 DPI)