BMJ Open

A NATIONWIDE LONGITUDINAL SURVEY AMONG FRENCH CANCER SURVIVORS TWO AND FIVE YEARS AFTER DIAGNOSIS: THE VICAN SURVEY.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005971
Article Type:	Protocol
Date Submitted by the Author:	26-Jun-2014
Complete List of Authors:	Bouhnik, Anne-Déborah; INSERM UMR912, Bendiane, Marc-Karim; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Cortaredona, Sebastien; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Sagaon-Teyssier, Luis; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Rey, Dominique; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Berenger, Cyril; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Seror, Valerie; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), Peretti-Watel, Patrick; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory,
Primary Subject Heading:	Public health
Secondary Subject Heading:	Oncology, Patient-centred medicine
Keywords:	ONCOLOGY, PUBLIC HEALTH, SOCIAL MEDICINE

A NATIONWIDE LONGITUDINAL SURVEY AMONG FRENCH CANCER SURVIVORS TWO AND FIVE YEARS AFTER DIAGNOSIS: THE VICAN SURVEY.

Anne-Deborah Bouhnik[1,2], Marc-Karim Bendiane[1,2,3], Sebastien Cortaredona[1,2,3], Luis Sagaon Teyssier[1,2,3], Dominique Rey[1,2,3], Cyril Berenger[1,2,3], Valerie Seror[1,2], Patrick Peretti-Watel[1,2,3]

1- INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), 13006, Marseille, France.

2- Aix Marseille University, UMR_S912, IRD, 13006, Marseille, France.

3- ORS PACA, Southeastern Health Regional Observatory, 13006, Marseille, France.

Corresponding author:

Anne-Deborah BOUHNIK, Institut Paoli-Calmettes, UMR912, 232 Boulevard Sainte Marguerite, 13009 Marseille, France. anne-deborah.bouhnik@inserm.fr

Acknowledgments: All the members of the VICAN Group: Thomas APARICIO, Emmanuel BABIN, François BECK, Robert BENAMOUZIG, Marc-Karim BEN DIANE, Cyril BERENGER, Dominique BESSETTE, Anne-Déborah BOUHNIK, Philippe-Jean BOUSQUET, Marie-Claude CABANEL-GICQUEL, Marianick CAVALLINI-LAMBERT, Michèle CHANTRY, Claire CHAUVET, Sébastien CORTAREDONA, Véronique DANGUY, Michel DORVAL, Jean-Baptiste HERBET, Laetitia HUIART, Xavier JOUTARD, Anne-Gaëlle LE CORROLLER-SORIANO, Julien MANCINI, Mégane MERESSE, Jean-François MORERE, Herman NABI, Alain PARAPONARIS, Patrick PERETTI-WATEL, Marie PREAU, Christel PROTIERE, Frédérique RETORNAZE, Dominique REY, Benoit RIANDEY, Luis SAGAON-TEYSSIER, Aurélia TISON, Valérie SEROR, Archana SINGH-MANOUX, Catherine THIEBLEMONT, Pierre VERGER.

ABSTRACT

Introduction: Today there is a growing need to expand research on cancer survivorship. This article describes the design and implementation of a national survey on French cancer survivors, the VICAN survey.

Method and Analysis: The target population included patients aged 18-82, diagnosed with a cancer between January and June 2010, and registered in one of the three main French Health Insurance Schemes. It was restricted to 12 tumour sites accounting for 88% of cancer incidence in France. The sampling was stratified with a non-proportional allocation, based on age (18-52 and 53-82 at diagnosis) and tumour site. Data collection includes telephone interviews with patients 2 and 5 years after diagnosis, a medical survey filled by the physician who initiated cancer treatment, and information collected from the national medico-administrative database on reimbursement data and hospital discharge records. First collection of data, 2 years after diagnosis, occurred between March and December 2012. Overall, 16,429 patients were initially contacted by mail. The final sample size was 4,349, with a global response rate of 43.7%. A weighting procedure was applied using the probabilities of selection in each stratum as well as other characteristics available for both respondents and non-respondents and correlated to participation: gender, age, socio-economic hardship, tumour site and cancer progression at time of the survey. Second data collection (5 years after diagnosis) will be conducted in 2015.

Ethics and Dissemination: The VICAN survey provides a powerful tool for public policies' evaluation and orientation in the short to medium term. It will also be a considerable dataset for behavioural and social sciences research in order to document "cancer survivorship" in the French context.

(266 words)

Keywords: cancer survivorship, national survey, longitudinal data, return to work, living conditions, quality of life.

Strengths and limitations of this study

- This longitudinal survey investigates various topics related to cancer survivorship among a large population-based national sample of 4,349 adults in France.
- This survey combines 3 sources of data: patient reported outcomes, medical records and medico-administrative databases.
- The use of medico-administrative databases to select participants resulted in a high level of ineligible patients and a high level of patients whose eligibility remained unknown, among those initially contacted.

BACKGROUND

Over the past decades, the incidence of most cancers has increased in developed countries [1-3]. At the same time, thanks to earlier diagnoses and more effective treatments, survival rates have also increased for most cancer [4-6]. As a result, the prevalence of individuals living with a history of cancer is steadily increasing. This has led to the need to expand research on cancer survivorship, and to consider it a very specific phase in the cancer journey [7-10]. In this article we consider that cancer survivorship begins after primary treatment and lasts until cancer recurrence or end of life [11], including many important aspects: follow-up treatment and second cancers, late side-effects of treatment, psychosocial adjustments and social relationships, return to work... [9 10 12 13].

In France, a first national survey on cancer survivors interviewed 2 years following diagnosis was conducted in 2004 [14]. This survey illustrated the enduring problems experienced by French cancer survivors who tried to stay within or return to the labour force, as well as the impact of socioeconomic status, clinical factors and perceived discrimination on job-loss risk [15]. It also showed that the return to work process was gender-specific [16]. Other analyses highlighted the strong impact of social inequalities on various outcomes, including quality of life, information-seeking behaviours, couple relationships and barriers to parenthood [17-20].

These results inspired the design of the 2009-2013 Cancer Plan [21], which also explicitly stipulated that the French national cancer institute (INCa) should support regular surveys to gather data on patients' living conditions after their cancer diagnosis. In accordance with this stipulation, the Research Unit SESSTIM (Economics and Social Sciences, Health Care System & Societies) was appointed to conduct a longitudinal survey on everyday life two and five years after a cancer diagnosis. This article aims to describe the design and implementation of this innovative and ambitious survey, which combines patients' self-reported data, information collected within their medical records, and administrative records for healthcare use.

METHODS

Definition of target population.

The survey targeted adult patients with cancer diagnosed between January and June 2010. As the active treatment phase does not usually last more than 12 months, targeted patients have experienced life after cancer for at least one year when first interviewed two years after diagnosis. People under 18 years-old at diagnosis were excluded from the survey for legal reasons. So too were those over 82 years-old at diagnosis for practical reasons: they represent 7% of cancer incidence [22], but they would have been aged >84 at the time of the survey, and telephone interviews can be quite difficult with such populations (they frequently live in institutions with no personal telephone line; they are prone to refuse telephone surveys, and hearing problems frequently complicate the interview [23 24]).

Health insurance is compulsory in France. All those treated for cancer are registered in the Long Duration Disease File of the National Health Insurance Fund (ALD file), with a code detailing the tumour site. For practical reasons, we restricted the survey to patients registered to one of the three main Health Insurance Schemes (CNAMTS for salaried workers, RSI for self-employed workers, MSA for farmers) covering more than 90% of the French population. Eligibility was restricted to French-speaking patients diagnosed with first malignant cancer and living in France for at least two years.

A sample stratified according to age and tumour site.

As our main objective was to investigate the barriers to and drivers of a return to work, we over-represented people aged <54 at diagnosis, as they were aged <56 at the time of the survey and therefore too young for retirement or early retirement schemes. Thus we defined two age strata, 18-52 and 53-82 at diagnosis, with a stratum weight of 50% for each.

We also restricted the survey to 12 tumour sites which accounted for 88% of global cancer incidence in France in 2012 [25]. Site selection depended on four criteria: global incidence, incidence by age (in coherence with age stratification), two-year survival rate and scientific interest (for example we planned to focus on lung cancer because of recent improvements in its survival). Selected tumour sites included cancers with good prognosis (breast, prostate

and thyroid cancers, melanoma), others with intermediate prognosis (colorectal, bladder, kidney, cervical, endometrial and upper aerodigestive tract cancers, Non-Hodgkin lymphoma) and one with poor prognosis (lung cancer) [26].

Sampling design

A simple random sampling was applied within each of the 24 strata (2 age ranges×12 tumour sites) using the ALD file. In order to over-represent people aged <54 at diagnosis and relatively rare tumour sites, we did not opt for proportional allocation. Sample sizes were determined *a priori* within each stratum in order to have enough statistical power to conduct analyses separately for certain tumour sites, for a global sample size of N=6,000 (see Table 1). We excluded prostate, bladder and endometrial cancers from the age stratum '18-52 at diagnosis', because these cancers have a much higher incidence among older people, and conversely, we excluded thyroid cancer from the stratum '53-82 at diagnosis', because its incidence sharply decreases after age 55 [22].

Table 1

For each stratum, we estimated the number of contacts necessary to achieve the target size using the response rate observed for the 2004 survey. We also took into account the difference in recruitment procedures between both surveys. For example, the desired sample size for the stratum 'breast cancer/18-52 at diagnosis' was N=800. In 2004, the observed response rate for this population was 59.7%. However, in 2004, patients were recruited by telephone by physicians from the National Health Insurance Fund, and the global response rate was 53.7%. Instead in 2012, we planned to recruit participants by postal mail (see *infra*), thus we expected a lower response rate. A survey conducted in 2007 (among patients with diabetes [27]) with the same recruitment procedure (a letter sent by the National Health Insurance Fund) reached a response rate of 45%. We expected a similar global response rate for the first data collection of VICAN. Therefore in 2012, for patients with breast cancer diagnosed at age 18-52, the expected response rate was 59.7%×45%/53.7%=50.1%. Consequently 1,597 patients had to be contacted to recruit 800 participants.

As we planned to recruit 6,000 participants, with an expected response rate of 45%, 13,333 (6,000/0.45) people registered in the ALD file should have been contacted. However, in the 2004 survey, a number of those contacted proved to be ineligible for various reasons, including inaccurate diagnosis encoded in the ALD file, and hospitalisations and deceases at the time of the survey. Accordingly, we slightly increased the number of scheduled contacts: 16,429 patients were contacted.

Data collection procedure.

Each selected patient received a letter inviting him/her to participate in the survey, sent by the National Health Insurance Fund. It did not mention the INCa or the word 'cancer', only the tumour site. For example, women with a diagnosis of breast cancer were asked to participate in a survey about their 'breast disease'. It was necessary because in the 2004 survey 7% of participants never used the word "cancer" during their interview [14]. The letter also mentioned that information would be collected from participants' medical records and administrative records. Those who agreed to participate had to send back signed informed consent letter. Those who did not respond were considered non-respondents. There was no dunning letter, as required by the French Commission on Individual Data Protection and Public Liberties (CNIL) that approved the study methodology.

Two years after diagnosis, participants were interviewed using the computer-assisted telephone interview (CATI) system. The questionnaire dealt with many topics (never mentioning the word "cancer"): socio-demographic background and socioeconomic status, circumstances of diagnosis, relationships with the healthcare system and health professionals, treatments received and perceived side effects, health-related quality of life (SF12 scale [28]), cancer-related fatigue (EORTC QLQ scale [29]), recent pain (DN4 and ID-Pain questionnaires [30 31]), occupational status since diagnosis, perceived discrimination, social support, diet and physical activity, alcohol and tobacco use, couple relationship and sexuality since diagnosis, and fertility preservation. A postal questionnaire was proposed to people with lung or upper aerodigestive tract cancer, as their condition could hamper their ability to respond orally.

For each participant, a medical survey was conducted with the physician who initiated cancer treatment, to collect detailed information regarding tumour histology (stage, grade, size) and treatments received. We also collected information from the national database SNIIR-AM, including reimbursement data (physicians and other health professionals' consultations, prescribed drugs) and hospital discharge records [32]. We also collected data measured at the residential area level to investigate spatial inequalities: socio-economic hardship indexes [33] [34], and measures of healthcare availability/accessibility (general practices, hospitals).

First data collection.

The first period of data collection, two years after diagnosis, occurred between March and December 2012. Telephone interviews lasted on average 40 minutes. Among those with lung or upper aerodigestive tract cancer, 68% asked for the postal questionnaire.

Among the 16,429 patients initially contacted, 50.5% returned the signed informed consent letter and 4% of the contact letters were returned because of inaccurate postal addresses. Patient eligibility was evaluated using three sources: a very brief questionnaire filled in by patients and returned with their consent, SNIIR-AM data and the medical survey. The proportion of eligible people was markedly lower than expected, especially among those aged 53-82 at diagnosis (55.1% *versus* 63.6% among patients aged 18-52 at diagnosis). The main reasons for ineligibility were: inaccurate diagnosis (51.5% of ineligible patients: benign or second cancers, or errors in ALD file regarding the tumour site), inappropriate delay between diagnosis and survey (21.7% of ineligible patients, in most cases late recording in the ALD file), and patient death before the survey (16.4%). In line with the recommendations of the American Association for Public Opinion Research, in order to compute a response rate, we assumed that the proportion of eligible people was identical among those who did not return the informed consent letter ('unknown eligibility') [35]. The resulting response rates were close to our expectations (42.8% for the age stratum '18-52 at diagnosis' and 44.5% for the age stratum '53-82 at diagnosis', an average of 43.7%), Due to the high proportion of ineligible people, the final sample size was only 4,349.

Table 2 details the sample's structure according to age and tumour site. Across the age×tumour site strata, the response rate varied between 37% (for women aged 53-82 at diagnosis of endometrial cancer) and 52% (18-52×Non-Hodgkin lymphoma).

Table 2

Regarding the medical survey, data collection took place between March 2012 and March 2013. After several reminders by phone and letters, it was completed for 87.7% of participants. SNIIR-AM data were collected for all participants.

Weighting procedures.

As we used a stratified random sampling design with a non-proportional allocation (see Table 1), we first computed sampling weights as reciprocals of the probabilities of selection in each stratum. Second, as we collected data from the SNIIR-AM file for all contacted patients, we had the opportunity to compare eligible respondents with eligible non-respondents.

Available information included gender, age, tumour site, and socio-economic hardship index [33]. We also expected that patients' health status was correlated to participation, thus we built an indicator of cancer progression thanks to SNIIR-AM data for every contacted patient. We considered that cancer was in progression for patients who met one of the following criteria: other cancer diagnosed since 2011, treatment with chemotherapy, radiotherapy or targeted therapy in 2012, admission to a palliative care unit in 2012, and decease.

The proportion of women was higher among respondents (52.8% *versus* 41.1% among non-respondents), who were also younger on average (60.4 years-old versus 64.7 for men), while non-respondents lived more frequently in areas with a high level of socio-economic hardship (see Table 3). Breast cancers were markedly over-represented among respondents (35.1% among respondents *versus* 22.4% among non-respondents), while upper aerodigestive tract and lung cancers were under-represented (all in all, 8.9% versus 17.2%). The proportion of cancers in progression was significantly higher among non-respondents (21.8% *versus* 17.4%).

Table 3

As gender, age, socio-economic hardship, tumour site and cancer progression all had a significant impact on participation, we adjusted initial weights for these five variables. The final weights were created using an iterative process (ranking ratio estimation). Thanks to the resulting weights, the marginal distribution for each of these variables was the same among all respondents (N=4,349), among eligible patients (N=4,876) and among patients whose eligibility status remained unknown (N=9,372).

Second data collection.

The National Health Insurance Fund will keep a matching file in order to propose participation to the same participants, five years after their cancer diagnosis. This second collection of data will be set up in 2015.

DISCUSSION

Objectives of the survey.

Before discussing these objectives, we must highlight that we collected two kinds of data: retrospective data (for example, concerning circumstances of diagnosis and treatments received) and data related to participants' current living conditions. Analyses using retrospective data may be influenced by a selection bias, as only "survivors" participated, except for tumour sites associated with a very good survival rate two years after diagnosis.

The aims of the VICAN survey were twofold. First, it was designed in close cooperation with the INCa, which is the official state agency in charge of coordinating public policies related to the fight against cancer. From this perspective, the VICAN survey will be a powerful tool for public policies' evaluation and orientation in the short to medium term. For example, in order to improve the way patients are informed of cancer diagnosis, the Cancer Plan 2009-2013 approved the generalisation of the "Diagnostic Disclosure Procedure" [21 36], whose context and content are precisely defined. Accordingly, specific questions were introduced in the first questionnaire to assess the implementation of this measure. Second, this survey was designed to stimulate social sciences research related to "cancer survivorship" in the French

context, as psychosocial issues predominate over medical ones during this specific phase of cancer journey [10 37].

The resort to medico-administrative databases.

Thanks to the specificities of the French healthcare system, we had the opportunity to use medico-administrative databases (ALD & SNIIR-AM files). These databases provide a convenient way to contact cancer survivors. In other countries, similar samples are built as subsamples of very large general population surveys (see for example [38]) or from cancer registries covering the whole territory [39], but such surveys/registries are not available in France. Moreover, these medico-administrative databases allowed us to target specific populations (especially regarding tumour site and time since diagnosis) and to collect data on both respondents and non-respondents (in order to detect and correct participation biases). These databases also provide detailed and reliable data regarding healthcare utilisation, while asking patients to self-report healthcare utilisation is both time-consuming and liable to recall bias.

However, using medico-administrative databases raises legal and technical issues that complicate the design of the survey. These databases also contain various kinds of inaccuracies, especially since they were not initially designed as research tools. In VICAN, these limits are illustrated by the relatively high level of ineligible patients among those who were initially contacted.

(2495 words)

Contributions: ADB, MKB, LST, DR, VS, PPW contributed to conception, design and management of the study. ADB, CB, MKB, PPW contributed to acquisition of data. ADB, MKB, LST, SC, CB, PPW contributed to data analysis. ADB, SC, PPW contributed to the draft of the manuscript. All authors critically revised successive drafts of the manuscript and approved the final version.

Competing interests: none

Funding: this study is funded by The National Institute of Cancer (INCa), "Contrat de recherche et développement n° 05-2011".

REFERENCES

1. Kanavos P. The rising burden of cancer in the developing world. *Ann Oncol* 2006;17 Suppl 8:viii15-viii23.
2. Arnold M, Karim-Kos HE, Coebergh JW, Byrnes G, Antilla A, Ferlay J, et al. Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. *Eur J Cancer* 2013.
3. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;49(6):1374-403.
4. Bosetti C, Bertuccio P, Malvezzi M, Levi F, Chatenoud L, Negri E, et al. Cancer mortality in Europe, 2005-2009, and an overview of trends since 1980. *Ann Oncol* 2013;24(10):2657-71.
5. Verdecchia A, Guzzinati S, Francisci S, De Angelis R, Bray F, Allemani C, et al. Survival trends in European cancer patients diagnosed from 1988 to 1999. *Eur J Cancer* 2009;45(6):1042-66.
6. Verdecchia A, Francisci S, Brenner H, Gatta G, Micheli A, Mangone L, et al. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. *Lancet Oncol* 2007;8(9):784-96.
7. McCabe MS. Looking beyond survival survivorship care is critical to the future of oncology. *Oncology (Williston Park)* 2007;21(3):398.
8. Rowland JH, Hewitt M, Ganz PA. Cancer survivorship: a new challenge in delivering quality cancer care. *J Clin Oncol* 2006;24(32):5101-4.
9. Ayanian JZ, Jacobsen PB. Enhancing research on cancer survivors. *J Clin Oncol* 2006;24(32):5149-53.
10. Grunfeld E. Looking beyond survival: how are we looking at survivorship? *J Clin Oncol* 2006;24(32):5166-9.
11. Hewitt M, Greenfield S, Stovall E. *From Cancer Patient to Cancer Survivor: Lost in Transition*. Washington, DC: The National Academies Press, 2005.
12. Mak AK, Chaidaroon S, Fan G, Thalib F. Unintended consequences: the social context of cancer survivors and work. *J Cancer Surviv* 2014.

13. Denlinger CS, Carlson RW, Are M, Baker KS, Davis E, Edge SB, et al. Survivorship: introduction and definition. Clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2014;12(1):34-45.
14. Le Corroller-Soriano A, Malavolti L, Mermillod C. *La vie deux ans après le diagnostic du cancer: une enquête en 2004 sur les conditions de vie des malades*. Paris: La documentation française, 2008.
15. Paraponaris A, Teyssier LS, Ventelou B. Job tenure and self-reported workplace discrimination for cancer survivors 2 years after diagnosis: does employment legislation matter? *Health Policy* 2010;98(2-3):144-55.
16. Marino P, Luis Sagaon T, Laetitia M, Anne-Gaelle le CS. Sex differences in the return-to-work process of cancer survivors 2 years after diagnosis: results from a large French population-based sample. *J Clin Oncol* 2013;31(10):1277-84.
17. Mancini J, Rey D, Preau M, Le Corroller-Soriano AG, Moatti JP. Barriers to procreational intentions among cancer survivors 2 years after diagnosis: a French national cross-sectional survey. *Psychooncology* 2011;20(1):12-8.
18. Preau M, Bouhnik AD, Rey D, Mancini J. Two years after cancer diagnosis, which couples become closer? *Eur J Cancer Care (Engl)* 2011;20(3):380-8.
19. Le Corroller-Soriano AG, Bouhnik AD, Preau M, Malavolti L, Julian-Reynier C, Auquier P, et al. Does cancer survivors' health-related quality of life depend on cancer type? Findings from a large French national sample 2 years after cancer diagnosis. *Eur J Cancer Care (Engl)* 2011;20(1):132-40.
20. Protiere C, Moumjid N, Bouhnik AD, Le Corroller Soriano AG, Moatti JP. Heterogeneity of cancer patient information-seeking behaviors. *Med Decis Making* 2012;32(2):362-75.
21. http://www.e-cancer.fr/en/cancer-plan-2009-2013, 2013.
22. Belot A, Grosclaude P, Bossard N, Jougla E, Benhamou E, Delafosse P, et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2008;56(3):159-75.
23. Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL, Waksberg J. *Telephone Survey Methodology*: John Wiley & Sons, 2001.
24. Beck F, Gautier A. *Baromètre cancer 2010*. Saint-Denis: INPES, 2012.
25. https://lesdonnees.e-cancer.fr/les-fiches-de-synthese/29-incidence-mortalite/38-ensemble-des-cancers/22-epidemiologie-des-cancers-en-france-metropolitaine-incidence-et-mortalite.html#ind2, 2013.
26. Mazeau-Woynar V, Cerf N. *Survie attendue des patients atteints de cancers en France: état des lieux*. Paris: Institut national du cancer, 2010.
27. Jourdan-Da Silva N. *Instauration d'une insulinothérapie dans le diabète de type 2: données épidémiologiques, étude entred 2001-2003*, 2005.
28. Burdine JN, Felix MR, Abel AL, Wiltraut CJ, Musselman YJ. The SF-12 as a population health measure: an exploratory examination of potential for application. *Health Serv Res* 2000;35(4):885-904.
29. Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J Clin Epidemiol* 1997;50(4):441-50.
30. Portenoy R. Development and testing of a neuropathic pain screening questionnaire: ID Pain. *Curr Med Res Opin* 2006;22(8):1555-65.

31. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005;114(1-2):29-36.
32. Lenormand F. Le système d'information de l'assurance maladie, le SNIIRAM et les échantillons de bénéficiaires. . *Courrier des Statistiques* 2005;113-114:33-51.
33. Carstairs V, Morris R. *Deprivation and Health in Scotland*. Aberdeen: Aberdeen University Press, 1991.
34. Rey G, Rican S, Jougla E. Mesure des inégalités de mortalité par cause de décès. Approche écologique à l'aide d'un indice de désavantage social. *Bulletin Epidémiolgique Hebdomadaire* 2011;8-9.
35. The American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition: AAPOR, 2011.
36. Dolbeault S, Bredart A. Cancer Diagnosis Disclosure: The French Experience. *New Challenges in Communication with Cancer Patients*. New-York: Springer, 2013.
37. Stanton AL. Psychosocial concerns and interventions for cancer survivors. *J Clin Oncol* 2006;24(32):5132-7.
38. Elliott J, Fallows A, Staetsky L, Smith PW, Foster CL, Maher EJ, et al. The health and well-being of cancer survivors in the UK: findings from a population-based survey. *Br J Cancer* 2011;105 Suppl 1:S11-20.
39. Smith T, Stein KD, Mehta CC, Kaw C, Kepner JL, Buskirk T, et al. The rationale, design, and implementation of the American Cancer Society's studies of cancer survivors. *Cancer* 2007;109(1):1-12.

Table 1. Sampling design: age at diagnosis & tumour sites strata (VICAN).

Age at diagnosis:	18-52 years old			53-82 years old		
Cancer type:	% in ALD file	expected % in sample	expected size	% in ALD file	expected % in sample	expected size
Breast cancer	40.6%	26.7%	800	17.1%	16.7%	500
Prostate cancer	___	___	___	39.2%	20.0%	600
Melanoma	7.3%	10.0%	300	2.1%	5.0%	150
Thyroid cancer	9.7%	10.0%	300	___	___	___
Colorectal cancer	8.7%	11.7%	350	14.2%	10.0%	300
Upper aerodigestive tract cancers	9.8%	10.0%	300	4.6%	6.7%	200
Bladder cancer	___	___	___	3.7%	6.7%	200
Kidney cancer	3.2%	5.0%	150	3.1%	5.0%	150
Non-Hodgkin lymphoma	5.3%	6.7%	200	3.5%	6.7%	200
Cervical cancer	4.3%	6.7%	200	2.5%	6.7%	200
Endometrial cancer	___	___	___	0.5%	3.3%	100
Lung cancer	11.1%	13.3%	400	9.5%	13.3%	400
Total:	100%	100%	3000	100%	100%	3000

Reading example: among patients aged 18-52 at diagnosis, for the 9 selected tumour sites, breast cancers represent 40.6% of patients registered in the ALD file in 2012. However, they were expected to represent only 26.7% in the corresponding sample (N=800).

Table 2. Final sample: age at diagnosis & tumour sites strata (VICAN).

	Age at diagnosis:		
Cancer type:	18-52	53-82	*Total*
Breast cancer	971	379	1,350
Prostate cancer	___	479	479
Melanoma	162	114	276
Thyroid cancer	181	___	181
Colorectal cancer	258	229	487
Upper aerodigestive tract cancers	153	131	284
Bladder cancer	___	143	143
Kidney cancer	108	110	218
Non-Hodgkin lymphoma	163	122	285
Cervical cancer	97	78	175
Endometrial cancer	___	75	75
Lung cancer	136	260	396
Total:	2,241	2,108	4,349

Table 3. Comparison between eligible respondents and eligible non-respondents (VICAN).

	respondents (N=4,349)	non-respondents (N=527)
	% column	
Gender:		
-men	47.2%	58.9%
-women	52.8%	41.1%***
Age: mean (SE)	60.4 (11.4)	64.7 (11.5)***
Social Deprivation Index:		
-<first quartile	20.3%	14.4%
- [1st-3rd quartiles]	33.4%	29.1%
->third quartile	46.3%	56.5%***
Cancer type:		
-breast cancer	35.1%	22.4%
-prostate cancer	24.8%	29.6%
-melanoma	3.1%	3.1%
-thyroid cancer	2.0%	1.4%
-colorectal cancer	11.8%	13.3%
-upper aerodigestive tract cancers	4.0%	9.1%
-bladder cancer	4.5%	4.8%
-kidney cancer	3.4%	3.6%
-non-Hodgkin lymphoma	3.2%	2.9%
-cervical cancer	1.3%	0.9%
-endometrial cancer	1.8%	0.8%
-lung cancer	4.9%	8.1%***
Cancer progression since diagnosis:		
-no	82.6%	78.2%
-yes	17.4%	21.8%***

***, **, *, ns: respectively statistically significant at $p<0.001$, $p<0.01$, $p<0.05$, not significant (Student's t-test for age, Pearson's χ^2 for other variables).

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	✓ 1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
	✓	(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	✓ 2	Explain the scientific background and rationale for the investigation being reported
Objectives	✓ 3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	✓ 4	Present key elements of study design early in the paper
Setting	✓ 5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	✓ 6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
not applicable		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	✓ 7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
not applicable	.	
Bias	✓ 9	Describe any efforts to address potential sources of bias
Study size	✓ 10	Explain how the study size was arrived at
Quantitative variables	✓ 11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	✓ 12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results			
Participants	✓	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
			(b) Give reasons for non-participation at each stage
			(c) Consider use of a flow diagram
Descriptive data	✓	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
			(b) Indicate number of participants with missing data for each variable of interest
			(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)
Outcome data		15*	*Cohort study*—Report numbers of outcome events or summary measures over time
not applicable.			*Case-control study*—Report numbers in each exposure category, or summary measures of exposure
			Cross-sectional study—Report numbers of outcome events or summary measures
Main results		16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
not applicable.			(*b*) Report category boundaries when continuous variables were categorized
			(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses		17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
not applicable.			
Discussion			
Key results	✓	18	Summarise key results with reference to study objectives
Limitations	✓	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	✓	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
not applicable. Generalisability		21	Discuss the generalisability (external validity) of the study results
Other information			
Funding	✓	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

LABOUR MARKET, PSYCHOSOCIAL OUTCOMES AND HEALTH CONDITIONS IN CANCER SURVIVORS: A NATIONWIDE LONGITUDINAL SURVEY TWO AND FIVE YEARS AFTER CANCER DIAGNOSIS (THE VICAN SURVEY).

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005971.R1
Article Type:	Protocol
Date Submitted by the Author:	17-Nov-2014
Complete List of Authors:	Bouhnik, Anne-Déborah; INSERM UMR912, Bendiane, Marc-Karim; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Cortaredona, Sebastien; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Sagaon-Teyssier, Luis; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Rey, Dominique; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Berenger, Cyril; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory, Seror, Valerie; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), Peretti-Watel, Patrick; INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), ; ORS PACA, Southeastern Health Regional Observatory,
Primary Subject Heading:	Public health
Secondary Subject Heading:	Oncology, Patient-centred medicine
Keywords:	ONCOLOGY, PUBLIC HEALTH, SOCIAL MEDICINE

LABOUR MARKET, PSYCHOSOCIAL OUTCOMES AND HEALTH CONDITIONS IN CANCER SURVIVORS: PROTOCOL FOR A NATIONWIDE LONGITUDINAL SURVEY TWO AND FIVE YEARS AFTER CANCER DIAGNOSIS (THE VICAN SURVEY).

Anne-Deborah Bouhnik[1,2], Marc-Karim Bendiane[1,2,3], Sebastien Cortaredona[1,2,3], Luis Sagaon Teyssier[1,2,3], Dominique Rey[1,2,3], Cyril Berenger[1,2,3], Valerie Seror[1,2], Patrick Peretti-Watel[1,2,3]

1- INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), 13006, Marseille, France.

2- Aix Marseille University, UMR_S912, IRD, 13006, Marseille, France.

3- ORS PACA, Southeastern Health Regional Observatory, 13006, Marseille, France.

Corresponding author:

Anne-Deborah BOUHNIK, Institut Paoli-Calmettes, UMR912, 232 Boulevard Sainte Marguerite, 13009 Marseille, France. anne-deborah.bouhnik@inserm.fr

Acknowledgments: All the members of the VICAN Group: Thomas APARICIO, Emmanuel BABIN, François BECK, Robert BENAMOUZIG, Marc-Karim BEN DIANE, Cyril BERENGER, Dominique BESSETTE, Anne-Déborah BOUHNIK, Philippe-Jean BOUSQUET, Marie-Claude CABANEL-GICQUEL, Marianick CAVALLINI-LAMBERT, Michèle CHANTRY, Claire CHAUVET, Sébastien CORTAREDONA, Véronique DANGUY, Michel DORVAL, Jean-Baptiste HERBET, Laetitia HUIART, Xavier JOUTARD, Anne-Gaëlle LE CORROLLER-SORIANO, Julien MANCINI, Mégane MERESSE, Jean-François MORERE, Herman NABI, Alain PARAPONARIS, Patrick PERETTI-WATEL, Marie PREAU, Christel PROTIERE, Frédérique RETORNAZE, Dominique REY, Benoit RIANDEY, Luis SAGAON-TEYSSIER, Aurélia TISON, Valérie SEROR, Archana SINGH-MANOUX, Catherine THIEBLEMONT, Pierre VERGER.

ABSTRACT

Introduction: Today a growing need exists for greater research into cancer survivorship, focusing on different spheres of the day-to-day life of diagnosed patients. This article describes the design and implementation of VICAN (“*VIe après le CANcer*”), a national survey on French cancer survivors.

Method and Analysis: The target population included patients aged 18-82, diagnosed with cancer between January and June 2010, and registered in one of the three main French Health Insurance Schemes. It was restricted to 12 tumour sites. Sampling was stratified using a non-proportional allocation, based on age at diagnosis (18-52 and 53-82) and tumour site. Data was collected from telephone interviews with patients 2 and 5 years after diagnosis, a medical survey completed by the physician who initiated cancer treatment, and information from the national medico-administrative database on reimbursement data and hospital discharge records. First data collection, 2 years after diagnosis, occurred between March and December 2012. Second data collection (5 years after diagnosis) will be conducted in 2015. Analyses will be conducted on various outcomes: quality-of-life, health status, psychosocial conditions with a particular focus on the impact of cancer diagnosis on labour market. The variety of measurements included in the survey will enable us to control for a wide range of factors.

Ethics and Dissemination: The methodology of the VICAN survey was approved by three national ethics commissions. Results of the study will be disseminated through national and international research conferences and in articles published in international peer-reviewed journals.

(239 words)

Keywords: cancer survivorship, national survey, longitudinal data, return to work, living conditions, quality of life.

Strengths and limitations of this study

This longitudinal survey investigates various topics related to cancer survivorship among a large population-based national sample of 4,349 adults in France.

The survey combines 3 sources of data: patient-reported outcomes, medical records and medico-administrative databases.

The use of medico-administrative databases to select participants from among all those initially contacted, resulted in a high number of ineligible patients and a high number of patients whose eligibility remained unknown.

BACKGROUND

Over the past decades, the incidence of most cancers has increased in developed countries [1-3]. At the same time, earlier diagnoses and more effective treatments have led to increased survival rates for most cancers [4-6]. As a result, the population of cancer survivors is steadily increasing. This has led to a growing number of studies on cancer survivorship, and to consider survivorship as a major stage in the continuum of care [7-10].

These studies have shown that many survivors face psychological, physical and social challenges that may impact their daily lives and their quality of life. Cancer therapies can create long-term health problems that may become permanent, such as fatigue[11 12], pain[13 14], lymphoedema[15], infertility[16-18], cognitive impairment[19], urinary disorders[20], and sexual dysfunction[21 22]. Cancer survivors are also at increased risk of developing a second cancer or treatment-related heart failure[23 24], years after the diagnosis of the initial cancer. Regarding the psychological effects, episodes of depression, anxiety, and distress may occur even long time after cancer diagnosis. Furthermore, depressive symptoms are often described in cancer survivors, with prevalence as high as 40% reported in those with lung cancer. However they are often under-diagnosed and under-treated [25-27]. The epidemiological evolution resulting from medical progress in screening and treatments has prompted the need to reconsider the position of people with cancer in terms of the disease's consequences, in particular at the occupational level. Indeed, the short- and long-term consequences of cancer treatment, as well as the initial diagnosis itself, can strongly influence not only work-based opportunities in terms of access into employment and a return-to-work, but also workplace activities focused on ensuring job tenure. In addition to the physical [28] and cognitive [29] limitations which may impair the participation of cancer survivors in the labour market, the role of negative relationships with co-workers [30 31] and employers [32] is an important one. Remaining in employment and the prospect of a return to work have both been identified as key aspects for cancer survivors' quality of life [33-35]. Several studies have underlined the need for comprehensive long-term care for cancer survivors[36 37] and emphasised the lack of data on the evolution of side effects of cancer treatments over time.

Despite pain being one of the symptoms which most affects patients' lives, the management of persistent pain is still often suboptimal[38]. To ensure a better quality of life for patients, to organize appropriate long-term follow-up for them, and to allow them to regain their place in society, it is necessary to acquire a greater understanding of the mid- and long-term physical and psychological consequences of the disease and their social impact. It was in this context that, the American Cancer Society's Studies of Cancer Survivors (SCS I-II) were initiated in 2007 in the United States [39]. In Europe, similar studies have been implemented, for example, the PROFILES registry in the Netherlands in 2011 [40]. Following the recommendations of the French national 2009-2013 Cancer Plan to financially support surveys collecting data on cancer survivors' living conditions, [41] the French National Cancer Institute (INCa) entrusted the implementation of VICAN ("*Vie après le CANcer*") - a national survey on French cancer survivors - to the INSERM UMR*

912 research unit.

Objectives of the VICAN survey.

The aim of the VICAN survey is to document the living conditions of adult cancer patients 2 and 5 years after cancer diagnosis. More specifically, the objectives are:

- First, to study the labour market outcomes.

  As an increasing number of people of working age are being diagnosed with cancer, growing importance is being attached to the workplace consequences of cancer [42 43 44]. However, some important aspects of this issue are often neglected and need further research [45]. In particular, the effect of cancer on an individual's employability needs to be disentangled from the effects of his/her socio-economic status. Integrating variables related to work characteristics will help us to understand the true effect of living with cancer on the individual survivor's economic situation. Important individual characteristics, such as economic status and psychosocial issues may either weaken or strengthen the effects that cancer has on job tenure and employability, and need to be documented. Furthermore,

the role of medical outcomes is often missing in related research studies. By simultaneously integrating variables related to an individual's economic situation, his/her social-economic characteristics, and medical data related to cancer, this survey will be able to shed some important light on the deleterious effects of cancer on working life at the individual level. The collection of data over a 5-year period after diagnosis will also allow us to describe the impact of cancer on professional trajectories and the transitions between different states in the workplace[44].

- Second, to determine the nature, prevalence, and temporality of factors that may negatively affect or improve the quality of life and daily life of cancer survivors, and to study their evolution at 2 and 5 years after cancer diagnosis. Health-related quality of life is a key element both in the evaluation of life after cancer diagnosis and in creating a balanced life for the individual. Accordingly, understanding the factors affecting long-term quality of life remains an important research issue[39 46]. Particular attention will be given in the survey to health status (treatment follow-up, management of treatment-related side-effects, comorbidities, cancer relapse or second cancer) and also to psychosocial conditions (lifestyle behaviours, perceived discrimination, family and social support). Relevant questions include for example: Are cancer sequelae diagnosed and treated well? What is the impact of long-term sequelae on people with cancer where the prognosis is very good? Do the changes in lifestyle behaviours impact on quality of life? What is the role of social inequalities?
- Third, to evaluate the physical, psychological and social needs of cancer survivors. For example, one of the questions to ask whether patients are satisfied with the information provided on treatment side-effects or on the risk of treatment-induced infertility?

- Fourth, to compare new data with results from a French survey performed in 2004[32 42 44 46-50].

A study 2 and 5 years after cancer diagnosis.

In this article we consider that cancer survivorship begins after primary treatment[51] . Therefore, we chose to implement the first part of the survey in patients 2 years after cancer diagnosis, effectively in the "recovery" phase, which follows the primary treatment phase. This choice allowed us to interview survivors who had cancer with intermediate or poor prognosis.

The second interview will occur in 2015, 5 years after cancer diagnosis, effectively at the end of the "early monitoring phase" (2-5 years after diagnosis), which is the period where the risk of relapse and of treatment side effects is greatest.

From the point of view of labour market outcomes, the choice of a survey 2 years after cancer diagnosis was based on the specificity of the social security system in France. State legislation provides considerable protection to workers and the impact of cancer diagnosis or of other chronic diseases on employability is quite different compared with many other countries, especially those where patients are confronted with a job-lock situation, whereby they are effectively tied to the same company in order to benefit from healthcare (e.g., in the United States). Indeed, in France and other countries with similar social security systems, little is known about the role played by sociodemographic, socioeconomic and clinical characteristics on the capacity of patients to retain their professional situation after diagnosis. Literature about other countries has demonstrated that the deleterious effect of cancer on professional trajectories begins to manifest itself at an early stage after diagnosis, and persists beyond the first 2 years [52]. This justifies the choice of interviewing the same individuals 2 and 5 years after diagnosis, as the information gathered may help us both to understand the situation French cancer survivors are confronted with in the labour market, and to analyse the extent to which the effects of cancer on labour market outcomes are irreversible.

This article aims to describe the design and implementation of this innovative and ambitious survey, which combines patients' self-reported data, information collected from their medical records, and administrative records for healthcare use.

METHODS

Definition of target population.

The survey targeted adult patients with cancer diagnosed between January and June 2010. As the active treatment phase does not usually last more than 12 months, targeted patients had experienced life after cancer for at least one year when first interviewed two years after diagnosis. People under 18 years old at diagnosis were excluded from the survey for legal reasons. Those over 82 years old at diagnosis were also excluded for practical reasons. Although the latter group represent 7% of cancer incidence in France [53], they would have been aged >84 at the time of first data collection, and telephone interviews with this age-group can be quite difficult for several reasons: they frequently live in institutions with no personal telephone line; they are prone to refuse telephone surveys; hearing problems frequently complicate the interview [54 55].

Health insurance is compulsory in France. All those treated for cancer are registered in the Long Duration Disease File of the National Health Insurance Fund (ALD file), with a code detailing the tumour site. For practical reasons, we restricted the survey to patients registered with one of the three main Health Insurance Schemes (CNAMTS for salaried workers, RSI for self-employed workers, MSA for farmers) which together cover more than 90% of the French population. Eligibility was restricted to French-speaking patients diagnosed with first malignant cancer and living in France for at least two years.

Sample stratified according to age and tumour site.

As our main objective was to investigate the barriers to and drivers of patients' return to work, we over-represented those aged <54 at diagnosis, as they were aged <56 at the time

of the survey and therefore too young for retirement or early retirement schemes. Accordingly, we defined two age strata - 18-52 and 53-82 at diagnosis - with a weight of 50% for each stratum.

We also restricted the survey to 12 tumour sites which accounted for 88% of global cancer incidence in France in 2012 [56]. Site selection depended on four criteria: global incidence, incidence by age (in line with our two age strata above), two-year survival rate and level of scientific interest (for example we planned to focus on lung cancer because of recent improvements in associated survival). Selected tumour sites included cancers with good prognosis (breast, prostate and thyroid cancers, melanoma), others with intermediate prognosis (colorectal, bladder, kidney, cervical, endometrial and upper aerodigestive tract cancers, Non-Hodgkin lymphoma), and one with poor prognosis (lung cancer) [57].

Sampling design.

A simple random sampling design was applied to each of the 24 strata (2 age ranges×12 tumour sites) using the ALD file. In order to over-represent both people aged <54 at diagnosis and relatively rare tumour sites, we did not opt for proportional allocation. Sample sizes were determined *a priori* within each stratum for a global sample size of N=6,000 (see Table 1). The objective was to have enough statistical power to conduct analyses separately for certain tumour sites, and to complete data collection within a reasonable period of time. Based on the experience of the survey carried out in 2004, which allowed us to use data collected for many topics[47], including employment[32 42 44], the chosen targeted sizes per tumour site and per age range seemed a good compromise. We excluded prostate, bladder and endometrial cancers from the age stratum '18-52 at diagnosis', because these cancers have a much higher incidence among older people. Conversely, we excluded thyroid cancer from the stratum '53-82 at diagnosis', because its incidence sharply decreases after age 55 [53].

Table 1

For each stratum, we estimated the number of contacts necessary to achieve the target size using the response rate observed for the 2004 survey. We also took into account the difference in recruitment procedures between both surveys. For example, the desired sample size for the stratum 'breast cancer/18-52 at diagnosis' was N=800 in the VICAN study. In 2004, the observed response rate for this population was 59.7%. However, in that survey, patients were recruited by telephone by physicians from the National Health Insurance Fund, and the global response rate was 53.7%. Instead in 2012, for the first data collection of VICAN, we planned to recruit participants by postal mail (see below). Accordingly, we expected a lower response rate. A survey conducted in 2007 (among patients with diabetes [58]) with the same recruitment procedure (a postal letter sent by the National Health Insurance Fund) had a response rate of 45%. We expected a similar global response rate for the first data collection of VICAN. Therefore in 2012, for patients with breast cancer diagnosed at age 18-52, the expected response rate was 59.7%×45%/53.7%=50.1%. Consequently 1,597 patients had to be contacted to recruit 800 participants.

As we planned to recruit 6,000 participants, with an expected response rate of 45%, 13,333 (6,000/0.45) people registered in the ALD file should have been contacted. However, in the 2004 survey, a number of those contacted proved to be ineligible for various reasons, including inaccurate diagnosis encoded in the ALD file, and hospitalisations and deceases at the time of the survey. Taking what had happened in the 2004 survey into account, we decided to slightly increase the number of scheduled contacts. In the end, 16,429 patients were contacted.

Data collection procedure.

Each selected patient received a letter inviting him/her to participate in the survey, sent by the National Health Insurance Fund. It did not mention the INCa or the word 'cancer', only the tumour site. For example, women with a diagnosis of breast cancer were asked to participate in a survey about their 'breast disease'. This approach was chosen because in the 2004 survey 7% of participants never used the word "cancer" during their interview [47]. The

letter also mentioned that information would be collected from participants' medical records and administrative records. Those who agreed to participate had to send back a signed informed consent letter. Those who did not respond were considered non-respondents. One dunning letter was sent. The study methodology was approved by three national ethics commissions: the CCTIRS (Comité Consultatif sur le Traitement de l'Information en Matière de Recherche dans le Domaine de la Santé, study registered under n°11-143), the ISP (Institute of Public Health, study registered under n°C11-63) and the CNIL (French Commission on Individual Data Protection and Public Liberties, study registered under n°911290).

In 2012, participants were interviewed using the computer-assisted telephone interview (CATI) system. A postal questionnaire was proposed to people with lung or upper aerodigestive tract cancer, as their condition could have hampered their ability to respond orally.

Data collected 2 years after diagnosis.

The CATI interview questionnaire dealt with many topics: socio-demographic background and socioeconomic status, circumstances of diagnosis, relationships with the healthcare system and health professionals, treatments received and perceived side effects. The word cancer was never mentioned. The questionnaire also included items related to perceived discrimination, social support, couple relationships, sexuality since diagnosis, and fertility preservation. Lifestyle-related outcomes such as diet and physical activity, alcohol and tobacco use were documented as well and several validated scales evaluated quality of life, fatigue and pain. Health-related quality of life was assessed using the French version of the SF12 scale [59]. Cancer-related fatigue was evaluated using the EORTC QLQ scale [60]. Pain was estimated using two validated scales: the DN4 and ID-Pain questionnaires [61] [62]. Moreover, since our survey focused on the impact of cancer on employment, a large part of the questionnaire was dedicated to this topic. Participants were asked about their working life during the study period (occupational status at the time of diagnosis and changes of status

over the study period), and their working conditions (type of job, work contract, work schedules, and income).They were also asked about the number and duration of periods of sick leave they had taken because of cancer. In addition, they were asked about perceived difficulty at work, and any work adjustments they had made or that had been proposed to them because of the disease.

For each participant, a medical survey was conducted with the physician who initiated cancer treatment, to collect detailed information regarding tumour histology (stage, grade, size) and treatments received. We also collected information from the national SNIIR-AM database, which includes financial reimbursement data (for physicians' and other health professionals' consultations, and for prescribed drugs) as well as hospital discharge records [63]. We also collected data measured at the residential area level to investigate spatial inequalities: socio-economic hardship indexes [64] [65], and measures of healthcare availability/accessibility (general practices, hospitals).

The patient and medical questionnaires are all available on the INCa website [66].

First data collection.

The first period of data collection, two years after diagnosis, occurred between March and December 2012. Telephone interviews lasted on average 40 minutes. Among those with lung or upper aerodigestive tract cancer who had the choice between a telephonic or postal interview, 68% asked for the latter.

Among the 16,429 patients initially contacted, 6,529 returned the signed informed consent form (see Figure 1). Patient eligibility was evaluated using three sources: a very brief questionnaire completed by patients and returned with their consent, SNIIR-AM data and the medical survey. In this context, among the 6,529 individuals who provided signed informed consent, 1653 were excluded because of non-eligibility. Similarly, among the 9,900 individuals who did not return the consent form, 1750 were identified as non-eligible. Consequently, of the 8,279 individuals whose eligibility or non-eligibility could be ascertained, only 58.9% were effectively eligible (55.1% and 63.6% among those aged 53-82 and aged

18-52 at diagnosis, respectively). The proportion of eligible people was markedly lower than expected, especially among those aged 53-82 at diagnosis.

Figure 1

The main reasons for ineligibility included inaccurate diagnosis (for 51.5% of ineligible patients: benign or second cancers, or errors in ALD file regarding the tumour site), inappropriate delay between diagnosis and survey (for 21.7% of ineligible patients: in most cases late recording in the ALD file), and patient death before the survey (for 16.4% of ineligible patients). In line with the recommendations of the American Association for Public Opinion Research, in order to compute a response rate, we assumed that the proportion of eligible people was identical among those who did not return the informed consent letter ('unknown eligibility') [67]. The resulting response rates were close to our expectations (42.8% for the age stratum '18-52 at diagnosis' and 44.5% for the age stratum '53-82 at diagnosis', providing an average of 43.7%). Due to the high proportion of ineligible people, the final sample size was only N=4,349.

Table 2 details the sample according to age and tumour site. Across the age×tumour site strata, the response rate varied between 37% (for women aged 53-82 at diagnosis of endometrial cancer) and 52% (18-52×Non-Hodgkin lymphoma).

Table 2

With respect to the medical survey, data collection took place between March 2012 and March 2013. After several reminders by phone and letters, it was completed for 87.7% of participants. SNIIR-AM data were collected for all participants.

Weighting procedures.

As we used a stratified random sampling design with non-proportional allocation (see Table 1), we first computed sampling weights as reciprocals of the probabilities of selection in each stratum. Second, as we collected data from the SNIIR-AM file for all contacted patients, we had the opportunity to compare eligible respondents with eligible non-respondents.

Available information included gender, age, tumour site, and socio-economic hardship index [64]. We also expected that patients' health status was correlated to participation. Accordingly, we built an indicator of cancer progression using SNIIR-AM data for every patient contacted. We considered that patients who met one of the following criteria had progressive cancer: second cancer diagnosed since 2011, treatment with chemotherapy, radiotherapy or targeted therapy in 2012, admission to a palliative care unit in 2012, and death.

The proportion of women was higher among respondents (52.8% *versus* 41.1% among non-respondents). Female respondents were also younger on average than their male counterparts (60.4 years-old versus 64.7), while non-respondents lived more frequently in areas with a high level of socio-economic hardship (see Table 3). Breast cancer was markedly over-represented among respondents (35.1% among respondents *versus* 22.4% among non-respondents), while upper aerodigestive tract and lung cancers were under-represented (overall 8.9% versus 17.2%). The proportion of individuals with progressive cancer was significantly higher among non-respondents (21.8% *versus* 17.4%).

Table 3

As gender, age, socio-economic hardship, tumour site and cancer progression all had a significant impact on participation, we adjusted initial weights for these five variables. The final weights were created using an iterative process (ranking ratio estimation). Thanks to the resulting weights, the marginal distribution for each of these variables was the same among all respondents (N=4,349), among eligible patients (N=4,876) and among patients whose eligibility status remained unknown (N=8,150).

Second data collection.

The National Health Insurance Fund will keep a matching file in order to propose participation to the same participants, five years after their cancer diagnosis. This second collection of data will occur in 2015.

Planned statistical analysis

Data analyses will be conducted using the SPSS software (PASW Statistics 18, version 18.0.3), Stata/SE software (version 12.1) or R (version 3.0.2). Cross-sectional and longitudinal analyses will be performed. For the former, multivariate linear or logistic regressions will be used, depending on the nature of the outcomes. For longitudinal analyses, linear or logistic mixed-model regressions will be used, depending on the nature of the outcomes, to account for repeated measurements. The variety of measurements included in the survey will enable us to control for a wide range of factors. A continuous-time Markov process model will be implemented to evaluate the impact of cancer diagnosis on mobility between the different states of the labour market (e.g., employment, unemployment, retirement, inactivity).

DISCUSSION

Objectives of the survey.

Before discussing the survey's objectives, we must highlight that we collected two kinds of data: retrospective data (e.g., concerning circumstances of diagnosis and treatments received), and data related to participants' current living conditions. Analyses using retrospective data may be influenced by selection bias, as only "survivors at 2 years" were interviewed and not everyone initially diagnosed with cancer. However this bias was certainly limited for tumour sites associated with a very good survival rate two years after diagnosis.

The VICAN survey has two main objectives. First, it was designed in close cooperation with the INCa, which is the official French state agency in charge of coordinating public policy related to the fight against cancer. From this perspective, the VICAN survey will be useful for public healthcare policy evaluation and orientation, in the short to medium term. For example, in order to improve the way patients are informed of cancer diagnosis, the Cancer Plan 2009-2013 approved the generalisation of the "Diagnostic Disclosure Procedure" [41 68], whose context and content are precisely defined. Accordingly, specific questions were introduced in the first questionnaire of the VICAN survey to assess the real world

implementation of this measure. Second, VICAN was designed to encourage social sciences research on "cancer survivorship" in the French context, as psychosocial issues dominate medical issues in this specific phase of the cancer trajectory [10 69].

Using medico-administrative databases.

Thanks to the specificities of the French healthcare system, we had the opportunity to use medico-administrative databases (ALD & SNIIR-AM files). These databases are now widely used for research purposes in many fields[70-73] and provide an effective way to contact cancer survivors. In other countries, similar samples have been created as subsamples of very large general population surveys [74] or from cancer registries covering the whole territory [39], but such surveys/registries are not available in France. Moreover, these medico-administrative databases allowed us to target specific populations (especially regarding tumour site and time since diagnosis) and to collect data on both respondents and non-respondents (in order to detect and correct for participation biases). These databases also provide detailed and reliable data regarding healthcare utilisation, unlike asking patients to self-report healthcare utilisation, which is both time-consuming and liable to recall bias. It is true however that many studies have shown that the effects of such biases on reported outcomes are minor [75-78].

Despite their value, the use of medico-administrative databases raises legal and technical issues that complicate the design of the survey. These databases also contain various kinds of inaccuracies, especially since they were not initially designed as research tools. In the VICAN survey, these limitations are illustrated by the relatively high level of ineligible patients among those who were initially contacted.

Ethics and Dissemination: The study methodology was approved by three national ethics commissions: the CCTIRS (Comité Consultatif sur le Traitement de l'Information en Matière de Recherche dans le Domaine de la Santé, study registered under n°11-143), the ISP (Institute of Public Health, study registered under n°C11-63) and the CNIL (French Commission on Individual Data Protection and Public Liberties, study registered under

n°911290). Confidentiality is assured for all participants with regard to any personal responses and information provided, as all data collected are anonymized. Results of the study will be disseminated through national and international research conferences and in articles published in international peer-reviewed journals.

(4176 words)

Contributions: ADB, MKB, LST, DR, VS, PPW contributed to the conception, design and management of the study. ADB, CB, MKB, PPW contributed to data collection. ADB, MKB, LST, SC, CB, PPW contributed to the data analysis. ADB, SC, LST, DR, PPW contributed to the drafting of the manuscript. All authors critically revised successive drafts of the manuscript and approved the final version.

Competing interests: none

Funding: This study was funded by The National Institute of Cancer (INCa), "Contrat de recherche et développement n° 05-2011".

REFERENCES

1. Kanavos P. The rising burden of cancer in the developing world. *Ann Oncol* 2006;17 Suppl 8:viii15-viii23.
2. Arnold M, Karim-Kos HE, Coebergh JW, Byrnes G, Antilla A, Ferlay J, et al. Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. *Eur J Cancer* 2013.
3. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;49(6):1374-403.
4. Bosetti C, Bertuccio P, Malvezzi M, Levi F, Chatenoud L, Negri E, et al. Cancer mortality in Europe, 2005-2009, and an overview of trends since 1980. *Ann Oncol* 2013;24(10):2657-71.

5. Verdecchia A, Guzzinati S, Francisci S, De Angelis R, Bray F, Allemani C, et al. Survival trends in European cancer patients diagnosed from 1988 to 1999. *Eur J Cancer* 2009;45(6):1042-66.
6. Verdecchia A, Francisci S, Brenner H, Gatta G, Micheli A, Mangone L, et al. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. *Lancet Oncol* 2007;8(9):784-96.
7. McCabe MS. Looking beyond survival survivorship care is critical to the future of oncology. *Oncology (Williston Park)* 2007;21(3):398.
8. Rowland JH, Hewitt M, Ganz PA. Cancer survivorship: a new challenge in delivering quality cancer care. *J Clin Oncol* 2006;24(32):5101-4.
9. Ayanian JZ, Jacobsen PB. Enhancing research on cancer survivors. *J Clin Oncol* 2006;24(32):5149-53.
10. Grunfeld E. Looking beyond survival: how are we looking at survivorship? *J Clin Oncol* 2006;24(32):5166-9.
11. Pertl MM, Quigley J, Hevey D. 'I'm not complaining because I'm alive': barriers to the emergence of a discourse of cancer-related fatigue. *Psychol Health* 2014;29(2):141-61.
12. Pachman DR, Barton DL, Swetz KM, Loprinzi CL. Troublesome symptoms in cancer survivors: fatigue, insomnia, neuropathy, and pain. *J Clin Oncol*;30(30):3687-96.
13. Marchettini P. More on pain semantics. *Eur J Pain* 2008;12(2):251; author reply 252.
14. Pachman DR, Barton DL, Swetz KM, Loprinzi CL. Troublesome symptoms in cancer survivors: fatigue, insomnia, neuropathy, and pain. *J Clin Oncol* 2012;30(30):3687-96.
15. Petrek JA, Senie RT, Peters M, Rosen PP. Lymphedema in a cohort of breast carcinoma survivors 20 years after diagnosis. *Cancer* 2001;92(6):1368-77.
16. Agarwal A, Allamaneni SS. Disruption of spermatogenesis by the cancer disease process. *J Natl Cancer Inst Monogr* 2005(34):9-12.
17. Armuand GM, Wettergren L, Rodriguez-Wallberg KA, Lampic C. Desire for children, difficulties achieving a pregnancy, and infertility distress 3 to 7 years after cancer diagnosis. *Support Care Cancer* 2014;22(10):2805-12.
18. Trivers KF, Fink AK, Partridge AH, Oktay K, Ginsburg ES, Li C, et al. Estimates of young breast cancer survivors at risk for infertility in the U.S. *Oncologist* 2014;19(8):814-22.
19. Rey D, Bouhnik AD, Mancini J, Bendiane MK, Seror V, Viens P. Self-reported cognitive impairment after breast cancer treatment in young women from the ELIPPSE40 cohort: the long-term impact of chemotherapy. *Breast J*;18(5):406-14.
20. Penson DF, McLerran D, Feng Z, Li L, Albertsen PC, Gilliland FD, et al. 5-year urinary and sexual outcomes after radical prostatectomy: results from the prostate cancer outcomes study. *J Urol* 2005;173(5):1701-5.
21. Rasmusson EM, Plantin L, Elmerstig E. 'Did they think I would understand all that on my own?' A questionnaire study about sexuality with Swedish cancer patients. *Eur J Cancer Care (Engl)* 2013;22(3):361-9.
22. Bober SL, Varela VS. Sexuality in adult cancer survivors: challenges and intervention. *J Clin Oncol* 2012;30(30):3712-9.
23. Bowles EJ, Wellman R, Feigelson HS, Onitilo AA, Freedman AN, Delate T, et al. Risk of heart failure in breast cancer patients after anthracycline and trastuzumab treatment: a retrospective cohort study. *J Natl Cancer Inst* 2012;104(17):1293-305.

24. Suter TM, Cook-Bruns N, Barton C. Cardiotoxicity associated with trastuzumab (Herceptin) therapy in the treatment of metastatic breast cancer. *Breast* 2004;13(3):173-83.
25. Alfano CM, Rowland JH. Recovery issues in cancer survivorship: a new challenge for supportive care. *Cancer J* 2006;12(5):432-43.
26. Sharpe M, Walker J, Holm Hansen C, Martin P, Symeonides S, Gourley C, et al. Integrated collabourative care for comorbid major depression in patients with cancer (SMaRT Oncology-2): a multicentre randomised controlled effectiveness trial. *Lancet* 2014;384(9948):1099-108.
27. Suppli NP, Johansen C, Christensen J, Kessing LV, Kroman N, Dalton SO. Increased Risk for Depression After Breast Cancer: A Nationwide Population-Based Cohort Study of Associated Factors in Denmark, 1998-2011. *J Clin Oncol* 2014.
28. Chirikos TN, Russell-Jacobs A, Jacobsen PB. Functional impairment and the economic consequences of female breast cancer. *Women Health* 2002;36(1):1-20.
29. Schagen SB, van Dam FS, Muller MJ, Boogerd W, Lindeboom J, Bruning PF. Cognitive deficits after postoperative adjuvant chemotherapy for breast carcinoma. *Cancer* 1999;85(3):640-50.
30. Greaves-Otte JG, Greaves J, Kruyt PM, van Leeuwen O, van der Wouden JC, van der Does E. Problems at social re-integration of long-term cancer survivors. *Eur J Cancer* 1991;27(2):178-81.
31. Maunsell E, Brisson C, Dubois L, Lauzier S, Fraser A. Work problems after breast cancer: an exploratory qualitative study. *Psychooncology* 1999;8(6):467-73.
32. Paraponaris A, Teyssier LS, Ventelou B. Job tenure and self-reported workplace discrimination for cancer survivors 2 years after diagnosis: does employment legislation matter? *Health Policy* 2010;98(2-3):144-55.
33. Kobayashi K, Morita S, Shimonagayoshi M, Kobayashi M, Fujiki Y, Uchida Y, et al. Effects of socioeconomic factors and cancer survivors' worries on their quality of life (QOL) in Japan. *Psychooncology* 2008;17(6):606-11.
34. Short PF, Vargo MM. Responding to employment concerns of cancer survivors. *J Clin Oncol* 2006;24(32):5138-41.
35. Taskila T, Lindbohm ML, Martikainen R, Lehto US, Hakanen J, Hietanen P. Cancer survivors' received and needed social support from their work place and the occupational health services. *Support Care Cancer* 2006;14(5):427-35.
36. Ganz PA, Hahn EE. Implementing a survivorship care plan for patients with breast cancer. *J Clin Oncol* 2008;26(5):759-67.
37. Grunfeld E. Optimizing follow-up after breast cancer treatment. *Curr Opin Obstet Gynecol* 2009;21(1):92-6.
38. Breivik H, Cherny N, Collett B, de Conno F, Filbet M, Foubert AJ, et al. Cancer-related pain: a pan-European survey of prevalence, treatment, and patient attitudes. *Ann Oncol* 2009;20(8):1420-33.
39. Smith T, Stein KD, Mehta CC, Kaw C, Kepner JL, Buskirk T, et al. The rationale, design, and implementation of the American Cancer Society's studies of cancer survivors. *Cancer* 2007;109(1):1-12.
40. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, et al. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an

infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. *Eur J Cancer* 2011;47(14):2188-94.

41. http://www.e-cancer.fr/en/cancer-plan-2009-2013, 2013.
42. Marino P, Luis Sagaon T, Laetitia M, Anne-Gaelle le CS. Sex differences in the return-to-work process of cancer survivors 2 years after diagnosis: results from a large French population-based sample. *J Clin Oncol* 2013;31(10):1277-84.
43. Mak AK, Chaidaroon S, Fan G, Thalib F. Unintended consequences: the social context of cancer survivors and work. *J Cancer Surviv* 2014.
44. Joutard X, Paraponaris A, Sagaon-Teyssier L, Ventelou B. Continuous-Time Markov Model for Transitions Between Employment and Non-Employment: The Impact of a Cancer Diagnosis. *Annals of Economics and Statistics* 2012;107/108.
45. de Boer AG, Taskila T, Ojajarvi A, van Dijk FJ, Verbeek JH. Cancer survivors and unemployment: a meta-analysis and meta-regression. *Jama* 2009;301(7):753-62.
46. Le Corroller-Soriano AG, Bouhnik AD, Preau M, Malavolti L, Julian-Reynier C, Auquier P, et al. Does cancer survivors' health-related quality of life depend on cancer type? Findings from a large French national sample 2 years after cancer diagnosis. *Eur J Cancer Care (Engl)* 2011;20(1):132-40.
47. Le Corroller-Soriano A, Malavolti L, Mermillod C. *La vie deux ans après le diagnostic du cancer: une enquête en 2004 sur les conditions de vie des malades*. Paris: La documentation française, 2008.
48. Mancini J, Rey D, Preau M, Le Corroller-Soriano AG, Moatti JP. Barriers to procreational intentions among cancer survivors 2 years after diagnosis: a French national cross-sectional survey. *Psychooncology* 2011;20(1):12-8.
49. Preau M, Bouhnik AD, Rey D, Mancini J. Two years after cancer diagnosis, which couples become closer? *Eur J Cancer Care (Engl)* 2011;20(3):380-8.
50. Protiere C, Moumjid N, Bouhnik AD, Le Corroller Soriano AG, Moatti JP. Heterogeneity of cancer patient information-seeking behaviors. *Med Decis Making* 2012;32(2):362-75.
51. Hewitt M, Greenfield S, Stovall E. *From Cancer Patient to Cancer Survivor: Lost in Transition*. Washington, DC: The National Academies Press, 2005.
52. Moran JR, Short PF, Hollenbeak CS. Long-term employment effects of surviving cancer. *J Health Econ*;30(3):505-14.
53. Belot A, Grosclaude P, Bossard N, Jougla E, Benhamou E, Delafosse P, et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2008;56(3):159-75.
54. Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL, Waksberg J. *Telephone Survey Methodology*: John Wiley & Sons, 2001.
55. Beck F, Gautier A. *Baromètre cancer 2010*. Saint-Denis: INPES, 2012.
56. https://lesdonnees.e-cancer.fr/les-fiches-de-synthese/29-incidence-mortalite/38-ensemble-des-cancers/22-epidemiologie-des-cancers-en-france-metropolitaine-incidence-et-mortalite.html#ind2, 2013.
57. Mazeau-Woynar V, Cerf N. *Survie attendue des patients atteints de cancers en France: état des lieux*. Paris: Institut national du cancer, 2010.
58. Jourdan-Da Silva N. *Instauration d'une insulinothérapie dans le diabète de type 2: données épidémiologiques, étude entred 2001-2003*, 2005.
59. Burdine JN, Felix MR, Abel AL, Wiltraut CJ, Musselman YJ. The SF-12 as a population health measure: an exploratory examination of potential for application. *Health Serv Res* 2000;35(4):885-904.

60. Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J Clin Epidemiol* 1997;50(4):441-50.
61. Portenoy R. Development and testing of a neuropathic pain screening questionnaire: ID Pain. *Curr Med Res Opin* 2006;22(8):1555-65.
62. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005;114(1-2):29-36.
63. Lenormand F. Le système d'information de l'assurance maladie, le SNIIRAM et les échantillons de bénéficiaires. . *Courrier des Statistiques* 2005;113-114:33-51.
64. Carstairs V, Morris R. *Deprivation and Health in Scotland*. Aberdeen: Aberdeen University Press, 1991.
65. Rey G, Rican S, Jougla E. Mesure des inégalités de mortalité par cause de décès. Approche écologique à l'aide d'un indice de désavantage social. *Bulletin Epidémiolgique Hebdomadaire* 2011;8-9.
66. http://www.e-cancer.fr/component/docman/doc_download/12164-la-vie-deux-ans-apres-un-diagnostic-de-cancer-de-lannonce-a-lapres-cancer-2014-annexes, 2014.
67. The American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition: AAPOR, 2011.
68. Dolbeault S, Bredart A. Cancer Diagnosis Disclosure: The French Experience. *New Challenges in Communication with Cancer Patients*. New-York: Springer, 2013.
69. Stanton AL. Psychosocial concerns and interventions for cancer survivors. *J Clin Oncol* 2006;24(32):5132-7.
70. Blotiere PO, Weill A, Ricordeau P, Alla F, Allemand H. Perforations and haemorrhages after colonoscopy in 2010: a study based on comprehensive French health insurance data (SNIIRAM). *Clin Res Hepatol Gastroenterol* 2014;38(1):112-7.
71. Fagot JP, Boutrelle A, Ricordeau P, Weill A, Allemand H. HPV vaccination in France: uptake, costs and issues for the National Health Insurance. *Vaccine* 2011;29(19):3610-6.
72. Neumann A, Maura G, Weill A, Ricordeau P, Alla F, Allemand H. Comparative effectiveness of rosuvastatin versus simvastatin in primary prevention among new users: a cohort study in the French national health insurance database. *Pharmacoepidemiol Drug Saf* 2014;23(3):240-50.
73. Santin G, Geoffroy B, Benezet L, Delezire P, Chatelot J, Sitta R, et al. In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse. *J Clin Epidemiol* 2014;67(6):722-30.
74. Elliott J, Fallows A, Staetsky L, Smith PW, Foster CL, Maher EJ, et al. The health and well-being of cancer survivors in the UK: findings from a population-based survey. *Br J Cancer* 2011;105 Suppl 1:S11-20.
75. Friedenreich CM, Howe GR, Miller AB. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. *Ann Epidemiol* 1991;1(5):439-53.
76. Friedenreich CM, Howe GR, Miller AB. The effect of recall bias on the association of calorie-providing nutrients and breast cancer. *Epidemiology* 1991;2(6):424-9.

77. Giovannucci E, Stampfer MJ, Colditz GA, Manson JE, Rosner BA, Longnecker MP, et al. Recall and selection bias in reporting past alcohol consumption among breast cancer cases. *Cancer Causes Control* 1993;4(5):441-8.
78. Vrijheid M, Deltour I, Krewski D, Sanchez M, Cardis E. The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. *J Expo Sci Environ Epidemiol* 2006;16(4):371-84.

Table 1. Sampling design: age at diagnosis & tumour sites strata (VICAN).

Age at diagnosis:	18-52 years old			53-82 years old		
Cancer type:	% in ALD file	expected % in sample	expected size	% in ALD file	expected % in sample	expected size
Breast cancer	40.6%	26.7%	800	17.1%	16.7%	500
Prostate cancer	___	___	___	39.2%	20.0%	600
Melanoma	7.3%	10.0%	300	2.1%	5.0%	150
Thyroid cancer	9.7%	10.0%	300	___	___	___
Colorectal cancer	8.7%	11.7%	350	14.2%	10.0%	300
Upper aerodigestive tract cancers	9.8%	10.0%	300	4.6%	6.7%	200
Bladder cancer	___	___	___	3.7%	6.7%	200
Kidney cancer	3.2%	5.0%	150	3.1%	5.0%	150
Non-Hodgkin lymphoma	5.3%	6.7%	200	3.5%	6.7%	200
Cervical cancer	4.3%	6.7%	200	2.5%	6.7%	200
Endometrial cancer	___	___	___	0.5%	3.3%	100
Lung cancer	11.1%	13.3%	400	9.5%	13.3%	400
Total:	100%	100%	3000	100%	100%	3000

Reading example: among patients aged 18-52 at diagnosis, for the 9 selected tumour sites, 40.6% of patients registered in the ALD file in 2012 had breast cancer. The expected figure was much lower in the sample: 26.7% (N=800).

Table 2. Final sample: age at diagnosis & tumour sites strata (VICAN).

Cancer type:	Age at diagnosis: 18-52	53-82	*Total*
Breast cancer	971	379	1,350
Prostate cancer	___	479	479
Melanoma	162	114	276
Thyroid cancer	181	___	181
Colorectal cancer	258	229	487
Upper aerodigestive tract cancers	153	131	284
Bladder cancer	___	143	143
Kidney cancer	108	110	218
Non-Hodgkin lymphoma	163	122	285
Cervical cancer	97	78	175
Endometrial cancer	___	75	75
Lung cancer	136	260	396
Total:	2,241	2,108	4,349

Table 3. Comparison between eligible respondents and eligible non-respondents (VICAN).

	respondents (N=4,349)	non-respondents (N=527)
	% column	
Gender:		
-men	47.2%	58.9%
-women	52.8%	41.1%***
Age: mean (SE)	60.4 (11.4)	64.7 (11.5)***
Social Deprivation Index:		
-<first quartile	20.3%	14.4%
- [1st-3rd quartiles]	33.4%	29.1%
->third quartile	46.3%	56.5%***
Cancer type:		
-breast cancer	35.1%	22.4%
-prostate cancer	24.8%	29.6%
-melanoma	3.1%	3.1%
-thyroid cancer	2.0%	1.4%
-colorectal cancer	11.8%	13.3%
-upper aerodigestive tract cancers	4.0%	9.1%
-bladder cancer	4.5%	4.8%
-kidney cancer	3.4%	3.6%
-non-Hodgkin lymphoma	3.2%	2.9%
-cervical cancer	1.3%	0.9%
-endometrial cancer	1.8%	0.8%
-lung cancer	4.9%	8.1%***
Cancer progression since diagnosis:	82.6%	78.2%
-no	17.4%	21.8%***
-yes		

***, **, *, ns: respectively statistically significant at p<0.001, p<0.01, p<0.05, not significant (Student's t-test for age, Pearson's χ^2 for other variables).

LABOUR MARKET, PSYCHOSOCIAL OUTCOMES AND HEALTH CONDITIONS IN CANCER SURVIVORS: A NATIONWIDE LONGITUDINAL SURVEY TWO AND FIVE YEARS AFTER CANCER DIAGNOSIS (THE VICAN SURVEY).

Anne-Deborah Bouhnik[1,2], Marc-Karim Bendiane[1,2,3], Sebastien Cortaredona[1,2,3], Luis Sagaon Teyssier[1,2,3], Dominique Rey[1,2,3], Cyril Berenger[1,2,3], Valerie Seror[1,2], Patrick Peretti-Watel[1,2,3]

1- INSERM, UMR912 "Economics and Social Sciences Applied to Health & Analysis of Medical Information" (SESSTIM), 13006, Marseille, France.

2- Aix Marseille University, UMR_S912, IRD, 13006, Marseille, France.

3- ORS PACA, Southeastern Health Regional Observatory, 13006, Marseille, France.

Corresponding author:

Anne-Deborah BOUHNIK, Institut Paoli-Calmettes, UMR912, 232 Boulevard Sainte Marguerite, 13009 Marseille, France. anne-deborah.bouhnik@inserm.fr

Acknowledgments: All the members of the VICAN Group: Thomas APARICIO, Emmanuel BABIN, François BECK, Robert BENAMOUZIG, Marc-Karim BEN DIANE, Cyril BERENGER, Dominique BESSETTE, Anne-Déborah BOUHNIK, Philippe-Jean BOUSQUET, Marie-Claude CABANEL-GICQUEL, Marianick CAVALLINI-LAMBERT, Michèle CHANTRY, Claire CHAUVET, Sébastien CORTAREDONA, Véronique DANGUY, Michel DORVAL, Jean-Baptiste HERBET, Laetitia HUIART, Xavier JOUTARD, Anne-Gaëlle LE CORROLLER-SORIANO, Julien MANCINI, Mégane MERESSE, Jean-François MORERE, Herman NABI, Alain PARAPONARIS, Patrick PERETTI-WATEL, Marie PREAU, Christel PROTIERE, Frédérique RETORNAZE, Dominique REY, Benoit RIANDEY, Luis SAGAON-TEYSSIER, Aurélia TISON, Valérie SEROR, Archana SINGH-MANOUX, Catherine THIEBLEMONT, Pierre VERGER.

ABSTRACT

Introduction: Today a growing need exists for greater research into cancer survivorship, focusing on different spheres of the day-to-day life of diagnosed patients. This article describes the design and implementation of VICAN, a national survey on French cancer survivors.

Method and Analysis: The target population included patients aged 18-82, diagnosed with cancer between January and June 2010, and registered in one of the three main French Health Insurance Schemes. It was restricted to 12 tumour sites accounting for 88% of cancer incidence in France. Sampling was stratified using a non-proportional allocation, based on age at diagnosis (18-52 and 53-82) and tumour site. Data was collected from telephone interviews with patients 2 and 5 years after diagnosis, a medical survey completed by the physician who initiated cancer treatment, and information from the national medico-administrative database on reimbursement data and hospital discharge records. First data collection, 2 years after diagnosis, occurred between March and December 2012. Overall, 16,429 patients were initially contacted by mail. The final sample size was 4,349, with a global response rate of 43.7%. A weighting procedure was applied using the probabilities of selection in each stratum and also the following characteristics (all correlated to participation) which were available for both respondents and non-respondents: gender, age, socio-economic status, tumour site and cancer progression at the time of the survey. Second data collection (5 years after diagnosis) will be conducted in 2015.

Ethics and Dissemination: The VICAN survey provides a powerful tool for public healthcare policy evaluation and orientation in the short to medium term. It will also generate a considerable dataset for behavioural and social sciences research in order to document "cancer survivorship" in the French context.

(278 words)

Keywords: cancer survivorship, national survey, longitudinal data, return to work, living conditions, quality of life.

Strengths and limitations of this study

This longitudinal survey investigates various topics related to cancer survivorship among a large population-based national sample of 4,349 adults in France.

The survey combines 3 sources of data: patient-reported outcomes, medical records and medico-administrative databases.

The use of medico-administrative databases to select participants from among all those initially contacted, resulted in a high number of ineligible patients and a high number of patients whose eligibility remained unknown.

BACKGROUND

Over the past decades, the incidence of most cancers has increased in developed countries [1-3]. At the same time, earlier diagnoses and more effective treatments have led to increased survival rates for most cancers [4-6]. As a result, the population of cancer survivors is steadily increasing. This has led to a growing number of studies on cancer survivorship, and to consider survivorship as a major stage in the continuum of care [7-10].

These studies have shown that many survivors face psychological, physical and social challenges that may impact their daily lives and their quality of life. Cancer therapies can create long-term health problems that may become permanent, such as fatigue[11 12], pain[13 14], lymphoedema[15], infertility[16-18], cognitive impairment[19], urinary disorders[20], and sexual dysfunction[21 22]. Cancer survivors are also at increased risk of developing a second cancer or treatment-related heart failure[23 24], years after the diagnosis of the initial cancer. Regarding the psychological effects, episodes of depression, anxiety, and distress may occur even long time after cancer diagnosis. Furthermore, depressive symptoms are often described in cancer survivors, with prevalence as high as 40% reported in those with lung cancer. However they are often under-diagnosed and under-treated [25-27]. The epidemiological evolution resulting from medical progress in screening and treatments has prompted the need to reconsider the position of people with cancer in terms of the disease's consequences, in particular at the occupational level. Indeed, the short- and long-term consequences of cancer treatment, as well as the initial diagnosis itself, can strongly influence not only work-based opportunities in terms of access into employment and a return-to-work, but also workplace activities focused on ensuring job tenure. In addition to the physical [28] and cognitive [29] limitations which may impair the participation of cancer survivors in the labour market, the role of negative relationships with co-workers [30 31] and employers [32] is an important one. Remaining in employment and the prospect of a return to work have both been identified as key aspects for cancer survivors' quality of life [33-35]. Several studies have underlined the need for comprehensive long-term care for cancer survivors[36 37] and emphasised the lack of data on the evolution of side effects of cancer treatments over time.

Despite pain being one of the symptoms which most affects patients' lives, the management of persistent pain is still often suboptimal[38]. To ensure a better quality of life for patients, to organize appropriate long-term follow-up for them, and to allow them to regain their place in society, it is necessary to acquire a greater understanding of the mid- and long-term physical and psychological consequences of the disease and their social impact. It was in this context that, the American Cancer Society's Studies of Cancer Survivors (SCS I-II) were initiated in 2007 in the United States [39]. In Europe, similar studies have been implemented, for example, the PROFILES registry in the Netherlands in 2011 [40]. Following the recommendations of the French national 2009-2013 Cancer Plan to financially support surveys collecting data on cancer survivors' living conditions, [41] the French National Cancer Institute (INCa) entrusted the implementation of VICAN - a national survey on French cancer survivors - to the INSERM UMR 912 research unit.

Objectives of the VICAN survey.

The aim of the VICAN survey is to document the living conditions of adult cancer patients 2 and 5 years after cancer diagnosis. More specifically, the objectives are:

- First, to study the labour market outcomes.

  As an increasing number of people of working age are being diagnosed with cancer, growing importance is being attached to the workplace consequences of cancer [42 43 44]. However, some important aspects of this issue are often neglected and need further research [45]. In particular, the effect of cancer on an individual's employability needs to be disentangled from the effects of his/her socio-economic status. Integrating variables related to work characteristics will help us to understand the true effect of living with cancer on the individual survivor's economic situation. Important individual characteristics, such as economic status and psychosocial issues may either weaken or strengthen the effects that cancer has on job tenure and employability, and need to be documented. Furthermore, the role of medical outcomes is often missing in related research studies. By

simultaneously integrating variables related to an individual's economic situation, his/her social-economic characteristics, and medical data related to cancer, this survey will be able to shed some important light on the deleterious effects of cancer on working life at the individual level. The collection of data over a 5-year period after diagnosis will also allow us to describe the impact of cancer on professional trajectories and the transitions between different states in the workplace[44].

- Second, to determine the nature, prevalence, and temporality of factors that may negatively affect or improve the quality of life and daily life of cancer survivors, and to study their evolution at 2 and 5 years after cancer diagnosis. Health-related quality of life is a key element both in the evaluation of life after cancer diagnosis and in creating a balanced life for the individual. Accordingly, understanding the factors affecting long-term quality of life remains an important research issue[39 46]. Particular attention will be given in the survey to health status (treatment follow-up, management of treatment-related side-effects, comorbidities, cancer relapse or second cancer) and also to psychosocial conditions (lifestyle behaviours, perceived discrimination, family and social support). Relevant questions include for example: Are cancer sequelae diagnosed and treated well? What is the impact of long-term sequelae on people with cancer where the prognosis is very good? Do the changes in lifestyle behaviours impact on quality of life? What is the role of social inequalities?
- Third, to evaluate the physical, psychological and social needs of cancer survivors. For example, one of the questions to ask whether patients are satisfied with the information provided on treatment side-effects or on the risk of treatment-induced infertility?

- Fourth, to compare new data with results from a French survey performed in 2004[32 42 44 46-50].

A study 2 and 5 years after cancer diagnosis.

In this article we consider that cancer survivorship begins after primary treatment[51] . Therefore, we chose to implement the first part of the survey in patients 2 years after cancer diagnosis, effectively in the "recovery" phase, which follows the primary treatment phase. This choice allowed us to interview survivors who had cancer with intermediate or poor prognosis.

The second interview will occur in 2015, 5 years after cancer diagnosis, effectively at the end of the "early monitoring phase" (2-5 years after diagnosis), which is the period where the risk of relapse and of treatment side effects is greatest.

From the point of view of labour market outcomes, the choice of a survey 2 years after cancer diagnosis was based on the specificity of the social security system in France. State legislation provides considerable protection to workers and the impact of cancer diagnosis or of other chronic diseases on employability is quite different compared with many other countries, especially those where patients are confronted with a job-lock situation, whereby they are effectively tied to the same company in order to benefit from healthcare (e.g., in the United States). Indeed, in France and other countries with similar social security systems, little is known about the role played by sociodemographic, socioeconomic and clinical characteristics on the capacity of patients to retain their professional situation after diagnosis. Literature about other countries has demonstrated that the deleterious effect of cancer on professional trajectories begins to manifest itself at an early stage after diagnosis, and persists beyond the first 2 years [52]. This justifies the choice of interviewing the same individuals 2 and 5 years after diagnosis, as the information gathered may help us both to understand the situation French cancer survivors are confronted with in the labour market, and to analyse the extent to which the effects of cancer on labour market outcomes are irreversible.

This article aims to describe the design and implementation of this innovative and ambitious survey, which combines patients' self-reported data, information collected from their medical records, and administrative records for healthcare use.

METHODS

Definition of target population.

The survey targeted adult patients with cancer diagnosed between January and June 2010. As the active treatment phase does not usually last more than 12 months, targeted patients had experienced life after cancer for at least one year when first interviewed two years after diagnosis. People under 18 years old at diagnosis were excluded from the survey for legal reasons. Those over 82 years old at diagnosis were also excluded for practical reasons. Although the latter group represent 7% of cancer incidence in France [53], they would have been aged >84 at the time of first data collection, and telephone interviews with this age-group can be quite difficult for several reasons: they frequently live in institutions with no personal telephone line; they are prone to refuse telephone surveys; hearing problems frequently complicate the interview [54 55].

Health insurance is compulsory in France. All those treated for cancer are registered in the Long Duration Disease File of the National Health Insurance Fund (ALD file), with a code detailing the tumour site. For practical reasons, we restricted the survey to patients registered with one of the three main Health Insurance Schemes (CNAMTS for salaried workers, RSI for self-employed workers, MSA for farmers) which together cover more than 90% of the French population. Eligibility was restricted to French-speaking patients diagnosed with first malignant cancer and living in France for at least two years.

Sample stratified according to age and tumour site.

As our main objective was to investigate the barriers to and drivers of patients' return to work, we over-represented those aged <54 at diagnosis, as they were aged <56 at the time

of the survey and therefore too young for retirement or early retirement schemes. Accordingly, we defined two age strata - 18-52 and 53-82 at diagnosis - with a weight of 50% for each stratum.

We also restricted the survey to 12 tumour sites which accounted for 88% of global cancer incidence in France in 2012 [56]. Site selection depended on four criteria: global incidence, incidence by age (in line with our two age strata above), two-year survival rate and level of scientific interest (for example we planned to focus on lung cancer because of recent improvements in associated survival). Selected tumour sites included cancers with good prognosis (breast, prostate and thyroid cancers, melanoma), others with intermediate prognosis (colorectal, bladder, kidney, cervical, endometrial and upper aerodigestive tract cancers, Non-Hodgkin lymphoma), and one with poor prognosis (lung cancer) [57].

Sampling design.

A simple random sampling design was applied to each of the 24 strata (2 age ranges×12 tumour sites) using the ALD file. In order to over-represent both people aged <54 at diagnosis and relatively rare tumour sites, we did not opt for proportional allocation. Sample sizes were determined *a priori* within each stratum for a global sample size of N=6,000 (see Table 1). The objective was to have enough statistical power to conduct analyses separately for certain tumour sites, and to complete data collection within a reasonable period of time. Based on the experience of the survey carried out in 2004, which allowed us to use data collected for many topics[47], including employment[32 42 44], the chosen targeted sizes per tumour site and per age range seemed a good compromise. We excluded prostate, bladder and endometrial cancers from the age stratum '18-52 at diagnosis', because these cancers have a much higher incidence among older people. Conversely, we excluded thyroid cancer from the stratum '53-82 at diagnosis', because its incidence sharply decreases after age 55 [53].

Table 1

For each stratum, we estimated the number of contacts necessary to achieve the target size using the response rate observed for the 2004 survey. We also took into account the difference in recruitment procedures between both surveys. For example, the desired sample size for the stratum 'breast cancer/18-52 at diagnosis' was N=800 in the VICAN study. In 2004, the observed response rate for this population was 59.7%. However, in that survey, patients were recruited by telephone by physicians from the National Health Insurance Fund, and the global response rate was 53.7%. Instead in 2012, for the first data collection of VICAN, we planned to recruit participants by postal mail (see below). Accordingly, we expected a lower response rate. A survey conducted in 2007 (among patients with diabetes [58]) with the same recruitment procedure (a postal letter sent by the National Health Insurance Fund) had a response rate of 45%. We expected a similar global response rate for the first data collection of VICAN. Therefore in 2012, for patients with breast cancer diagnosed at age 18-52, the expected response rate was 59.7%×45%/53.7%=50.1%. Consequently 1,597 patients had to be contacted to recruit 800 participants.

As we planned to recruit 6,000 participants, with an expected response rate of 45%, 13,333 (6,000/0.45) people registered in the ALD file should have been contacted. However, in the 2004 survey, a number of those contacted proved to be ineligible for various reasons, including inaccurate diagnosis encoded in the ALD file, and hospitalisations and deceases at the time of the survey. Taking what had happened in the 2004 survey into account, we decided to slightly increase the number of scheduled contacts. In the end, 16,429 patients were contacted.

Data collection procedure.

Each selected patient received a letter inviting him/her to participate in the survey, sent by the National Health Insurance Fund. It did not mention the INCa or the word 'cancer', only the tumour site. For example, women with a diagnosis of breast cancer were asked to participate in a survey about their 'breast disease'. This approach was chosen because in the 2004 survey 7% of participants never used the word "cancer" during their interview [47]. The

letter also mentioned that information would be collected from participants' medical records and administrative records. Those who agreed to participate had to send back a signed informed consent letter. Those who did not respond were considered non-respondents. No dunning letter was sent. The study methodology was approved by three national ethics commissions: the CCTIRS (Comité Consultatif sur le Traitement de l'Information en Matière de Recherche dans le Domaine de la Santé, study registered under n°11-143), the ISP (Institute of Public Health, study registered under n°C11-63) and the CNIL (French Commission on Individual Data Protection and Public Liberties, study registered under n°911290).

In 2012, participants were interviewed using the computer-assisted telephone interview (CATI) system. A postal questionnaire was proposed to people with lung or upper aerodigestive tract cancer, as their condition could have hampered their ability to respond orally.

Data collected 2 years after diagnosis.

The CATI interview questionnaire dealt with many topics: socio-demographic background and socioeconomic status, circumstances of diagnosis, relationships with the healthcare system and health professionals, treatments received and perceived side effects. The word cancer was never mentioned. The questionnaire also included items related to perceived discrimination, social support, couple relationships, sexuality since diagnosis, and fertility preservation. Lifestyle-related outcomes such as diet and physical activity, alcohol and tobacco use were documented as well and several validated scales evaluated quality of life, fatigue and pain. Health-related quality of life was assessed using the French version of the SF12 scale [59]. Cancer-related fatigue was evaluated using the EORTC QLQ scale [60]. Pain was estimated using two validated scales: the DN4 and ID-Pain questionnaires [61] [62]. Moreover, since our survey focused on the impact of cancer on employment, a large part of the questionnaire was dedicated to this topic. Participants were asked about their working life during the study period (occupational status at the time of diagnosis and changes of status

over the study period), and their working conditions (type of job, work contract, work schedules, and income).They were also asked about the number and duration of periods of sick leave they had taken because of cancer. In addition, they were asked about perceived difficulty at work, and any work adjustments they had made or that had been proposed to them because of the disease.

For each participant, a medical survey was conducted with the physician who initiated cancer treatment, to collect detailed information regarding tumour histology (stage, grade, size) and treatments received. We also collected information from the national SNIIR-AM database, which includes financial reimbursement data (for physicians' and other health professionals' consultations, and for prescribed drugs) as well as hospital discharge records [63]. We also collected data measured at the residential area level to investigate spatial inequalities: socio-economic hardship indexes [64] [65], and measures of healthcare availability/accessibility (general practices, hospitals).

The patient and medical questionnaires are all available on the INCa website [66].

First data collection.

The first period of data collection, two years after diagnosis, occurred between March and December 2012. Telephone interviews lasted on average 40 minutes. Among those with lung or upper aerodigestive tract cancer who had the choice between a telephonic or postal interview, 68% asked for the latter.

Among the 16,429 patients initially contacted, 6,529 returned the signed informed consent form (see Figure 1). Patient eligibility was evaluated using three sources: a very brief questionnaire completed by patients and returned with their consent, SNIIR-AM data and the medical survey. In this context, among the 6,529 individuals who provided signed informed consent, 1653 were excluded because of non-eligibility. Similarly, among the 9,900 individuals who did not return the consent form, 1750 were identified as non-eligible. Consequently, of the 8,279 individuals whose eligibility or non-eligibility could be ascertained, only 58.9% were effectively eligible (55.1% and 63.6% among those aged 53-82 and aged

18-52 at diagnosis, respectively). The proportion of eligible people was markedly lower than expected, especially among those aged 53-82 at diagnosis.

Figure 1: Sample selection pathway (VICAN).

The main reasons for ineligibility included inaccurate diagnosis (for 51.5% of ineligible patients: benign or second cancers, or errors in ALD file regarding the tumour site), inappropriate delay between diagnosis and survey (for 21.7% of ineligible patients: in most cases late recording in the ALD file), and patient death before the survey (for 16.4% of ineligible patients). In line with the recommendations of the American Association for Public Opinion Research, in order to compute a response rate, we assumed that the proportion of eligible people was identical among those who did not return the informed consent letter ('unknown eligibility') [67]. The resulting response rates were close to our expectations (42.8% for the age stratum '18-52 at diagnosis' and 44.5% for the age stratum '53-82 at diagnosis', providing an average of 43.7%). Due to the high proportion of ineligible people, the final sample size was only N=4,349.

Table 2 details the sample according to age and tumour site. Across the age×tumour site strata, the response rate varied between 37% (for women aged 53-82 at diagnosis of endometrial cancer) and 52% (18-52×Non-Hodgkin lymphoma).

Table 2

With respect to the medical survey, data collection took place between March 2012 and March 2013. After several reminders by phone and letters, it was completed for 87.7% of participants. SNIIR-AM data were collected for all participants.

Weighting procedures.

As we used a stratified random sampling design with non-proportional allocation (see Table 1), we first computed sampling weights as reciprocals of the probabilities of selection in each stratum. Second, as we collected data from the SNIIR-AM file for all contacted patients, we had the opportunity to compare eligible respondents with eligible non-respondents.

Available information included gender, age, tumour site, and socio-economic hardship index [64]. We also expected that patients' health status was correlated to participation. Accordingly, we built an indicator of cancer progression using SNIIR-AM data for every patient contacted. We considered that patients who met one of the following criteria had progressive cancer: second cancer diagnosed since 2011, treatment with chemotherapy, radiotherapy or targeted therapy in 2012, admission to a palliative care unit in 2012, and death.

The proportion of women was higher among respondents (52.8% *versus* 41.1% among non-respondents). Female respondents were also younger on average than their male counterparts (60.4 years-old versus 64.7), while non-respondents lived more frequently in areas with a high level of socio-economic hardship (see Table 3). Breast cancer was markedly over-represented among respondents (35.1% among respondents *versus* 22.4% among non-respondents), while upper aerodigestive tract and lung cancers were under-represented (overall 8.9% versus 17.2%). The proportion of individuals with progressive cancer was significantly higher among non-respondents (21.8% *versus* 17.4%).

Table 3

As gender, age, socio-economic hardship, tumour site and cancer progression all had a significant impact on participation, we adjusted initial weights for these five variables. The final weights were created using an iterative process (ranking ratio estimation). Thanks to the resulting weights, the marginal distribution for each of these variables was the same among all respondents (N=4,349), among eligible patients (N=4,876) and among patients whose eligibility status remained unknown (N=8,150).

Second data collection.

The National Health Insurance Fund will keep a matching file in order to propose participation to the same participants, five years after their cancer diagnosis. This second collection of data will occur in 2015.

Planned statistical analysis

Data analyses will be conducted using the SPSS software (PASW Statistics 18, version 18.0.3), Stata/SE software (version 12.1) or R (version 3.0.2). Cross-sectional and longitudinal analyses will be performed. For the former, multivariate linear or logistic regressions will be used, depending on the nature of the outcomes. For longitudinal analyses, linear or logistic mixed-model regressions will be used, depending on the nature of the outcomes, to account for repeated measurements. The variety of measurements included in the survey will enable us to control for a wide range of factors. A continuous-time Markov process model will be implemented to evaluate the impact of cancer diagnosis on mobility between the different states of the labour market (e.g., employment, unemployment, retirement, inactivity).

DISCUSSION

Objectives of the survey.

Before discussing the survey's objectives, we must highlight that we collected two kinds of data: retrospective data (e.g., concerning circumstances of diagnosis and treatments received), and data related to participants' current living conditions. Analyses using retrospective data may be influenced by selection bias, as only "survivors at 2 years" were interviewed and not everyone initially diagnosed with cancer. However this bias was certainly limited for tumour sites associated with a very good survival rate two years after diagnosis.

The VICAN survey has two main objectives. First, it was designed in close cooperation with the INCa, which is the official French state agency in charge of coordinating public policy related to the fight against cancer. From this perspective, the VICAN survey will be a powerful tool for public healthcare policy evaluation and orientation, in the short to medium term. For example, in order to improve the way patients are informed of cancer diagnosis, the Cancer Plan 2009-2013 approved the generalisation of the "Diagnostic Disclosure Procedure" [41 68], whose context and content are precisely defined. Accordingly, specific questions were introduced in the first questionnaire of the VICAN survey to assess the real

world implementation of this measure. Second, VICAN was designed to encourage social sciences research on “cancer survivorship” in the French context, as psychosocial issues dominate medical issues in this specific phase of the cancer trajectory [10 69].

Using medico-administrative databases.

Thanks to the specificities of the French healthcare system, we had the opportunity to use medico-administrative databases (ALD & SNIIR-AM files). These databases are now widely used for research purposes in many fields[70-73] and provide an effective way to contact cancer survivors. In other countries, similar samples have been created as subsamples of very large general population surveys [74] or from cancer registries covering the whole territory [39], but such surveys/registries are not available in France. Moreover, these medico-administrative databases allowed us to target specific populations (especially regarding tumour site and time since diagnosis) and to collect data on both respondents and non-respondents (in order to detect and correct for participation biases). These databases also provide detailed and reliable data regarding healthcare utilisation, unlike asking patients to self-report healthcare utilisation, which is both time-consuming and liable to recall bias. It is true however that many studies have shown that the effects of such biases on reported outcomes are minor [75-78].

Despite their value, the use of medico-administrative databases raises legal and technical issues that complicate the design of the survey. These databases also contain various kinds of inaccuracies, especially since they were not initially designed as research tools. In the VICAN survey, these limitations are illustrated by the relatively high level of ineligible patients among those who were initially contacted.

(4074 words)

Contributions: ADB, MKB, LST, DR, VS, PPW contributed to the conception, design and management of the study. ADB, CB, MKB, PPW contributed to data collection. ADB, MKB, LST, SC, CB, PPW contributed to the data analysis. ADB, SC, LST, DR, PPW contributed to

the drafting of the manuscript. All authors critically revised successive drafts of the manuscript and approved the final version.

Competing interests: none

Funding: This study was funded by The National Institute of Cancer (INCa), "Contrat de recherche et développement n° 05-2011".

REFERENCES

1. Kanavos P. The rising burden of cancer in the developing world. *Ann Oncol* 2006;17 Suppl 8:viii15-viii23.
2. Arnold M, Karim-Kos HE, Coebergh JW, Byrnes G, Antilla A, Ferlay J, et al. Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. *Eur J Cancer* 2013.
3. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;49(6):1374-403.
4. Bosetti C, Bertuccio P, Malvezzi M, Levi F, Chatenoud L, Negri E, et al. Cancer mortality in Europe, 2005-2009, and an overview of trends since 1980. *Ann Oncol* 2013;24(10):2657-71.
5. Verdecchia A, Guzzinati S, Francisci S, De Angelis R, Bray F, Allemani C, et al. Survival trends in European cancer patients diagnosed from 1988 to 1999. *Eur J Cancer* 2009;45(6):1042-66.
6. Verdecchia A, Francisci S, Brenner H, Gatta G, Micheli A, Mangone L, et al. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. *Lancet Oncol* 2007;8(9):784-96.
7. McCabe MS. Looking beyond survival survivorship care is critical to the future of oncology. *Oncology (Williston Park)* 2007;21(3):398.
8. Rowland JH, Hewitt M, Ganz PA. Cancer survivorship: a new challenge in delivering quality cancer care. *J Clin Oncol* 2006;24(32):5101-4.
9. Ayanian JZ, Jacobsen PB. Enhancing research on cancer survivors. *J Clin Oncol* 2006;24(32):5149-53.
10. Grunfeld E. Looking beyond survival: how are we looking at survivorship? *J Clin Oncol* 2006;24(32):5166-9.
11. Pertl MM, Quigley J, Hevey D. 'I'm not complaining because I'm alive': barriers to the emergence of a discourse of cancer-related fatigue. *Psychol Health* 2014;29(2):141-61.
12. Pachman DR, Barton DL, Swetz KM, Loprinzi CL. Troublesome symptoms in cancer survivors: fatigue, insomnia, neuropathy, and pain. *J Clin Oncol*;30(30):3687-96.

13. Marchettini P. More on pain semantics. *Eur J Pain* 2008;12(2):251; author reply 252.
14. Pachman DR, Barton DL, Swetz KM, Loprinzi CL. Troublesome symptoms in cancer survivors: fatigue, insomnia, neuropathy, and pain. *J Clin Oncol* 2012;30(30):3687-96.
15. Petrek JA, Senie RT, Peters M, Rosen PP. Lymphedema in a cohort of breast carcinoma survivors 20 years after diagnosis. *Cancer* 2001;92(6):1368-77.
16. Agarwal A, Allamaneni SS. Disruption of spermatogenesis by the cancer disease process. *J Natl Cancer Inst Monogr* 2005(34):9-12.
17. Armuand GM, Wettergren L, Rodriguez-Wallberg KA, Lampic C. Desire for children, difficulties achieving a pregnancy, and infertility distress 3 to 7 years after cancer diagnosis. *Support Care Cancer* 2014;22(10):2805-12.
18. Trivers KF, Fink AK, Partridge AH, Oktay K, Ginsburg ES, Li C, et al. Estimates of young breast cancer survivors at risk for infertility in the U.S. *Oncologist* 2014;19(8):814-22.
19. Rey D, Bouhnik AD, Mancini J, Bendiane MK, Seror V, Viens P. Self-reported cognitive impairment after breast cancer treatment in young women from the ELIPPSE40 cohort: the long-term impact of chemotherapy. *Breast J*;18(5):406-14.
20. Penson DF, McLerran D, Feng Z, Li L, Albertsen PC, Gilliland FD, et al. 5-year urinary and sexual outcomes after radical prostatectomy: results from the prostate cancer outcomes study. *J Urol* 2005;173(5):1701-5.
21. Rasmusson EM, Plantin L, Elmerstig E. 'Did they think I would understand all that on my own?' A questionnaire study about sexuality with Swedish cancer patients. *Eur J Cancer Care (Engl)* 2013;22(3):361-9.
22. Bober SL, Varela VS. Sexuality in adult cancer survivors: challenges and intervention. *J Clin Oncol* 2012;30(30):3712-9.
23. Bowles EJ, Wellman R, Feigelson HS, Onitilo AA, Freedman AN, Delate T, et al. Risk of heart failure in breast cancer patients after anthracycline and trastuzumab treatment: a retrospective cohort study. *J Natl Cancer Inst* 2012;104(17):1293-305.
24. Suter TM, Cook-Bruns N, Barton C. Cardiotoxicity associated with trastuzumab (Herceptin) therapy in the treatment of metastatic breast cancer. *Breast* 2004;13(3):173-83.
25. Alfano CM, Rowland JH. Recovery issues in cancer survivorship: a new challenge for supportive care. *Cancer J* 2006;12(5):432-43.
26. Sharpe M, Walker J, Holm Hansen C, Martin P, Symeonides S, Gourley C, et al. Integrated collabourative care for comorbid major depression in patients with cancer (SMaRT Oncology-2): a multicentre randomised controlled effectiveness trial. *Lancet* 2014;384(9948):1099-108.
27. Suppli NP, Johansen C, Christensen J, Kessing LV, Kroman N, Dalton SO. Increased Risk for Depression After Breast Cancer: A Nationwide Population-Based Cohort Study of Associated Factors in Denmark, 1998-2011. *J Clin Oncol* 2014.
28. Chirikos TN, Russell-Jacobs A, Jacobsen PB. Functional impairment and the economic consequences of female breast cancer. *Women Health* 2002;36(1):1-20.
29. Schagen SB, van Dam FS, Muller MJ, Boogerd W, Lindeboom J, Bruning PF. Cognitive deficits after postoperative adjuvant chemotherapy for breast carcinoma. *Cancer* 1999;85(3):640-50.
30. Greaves-Otte JG, Greaves J, Kruyt PM, van Leeuwen O, van der Wouden JC, van der Does E. Problems at social re-integration of long-term cancer survivors. *Eur J Cancer* 1991;27(2):178-81.

31. Maunsell E, Brisson C, Dubois L, Lauzier S, Fraser A. Work problems after breast cancer: an exploratory qualitative study. *Psychooncology* 1999;8(6):467-73.
32. Paraponaris A, Teyssier LS, Ventelou B. Job tenure and self-reported workplace discrimination for cancer survivors 2 years after diagnosis: does employment legislation matter? *Health Policy* 2010;98(2-3):144-55.
33. Kobayashi K, Morita S, Shimonagayoshi M, Kobayashi M, Fujiki Y, Uchida Y, et al. Effects of socioeconomic factors and cancer survivors' worries on their quality of life (QOL) in Japan. *Psychooncology* 2008;17(6):606-11.
34. Short PF, Vargo MM. Responding to employment concerns of cancer survivors. *J Clin Oncol* 2006;24(32):5138-41.
35. Taskila T, Lindbohm ML, Martikainen R, Lehto US, Hakanen J, Hietanen P. Cancer survivors' received and needed social support from their work place and the occupational health services. *Support Care Cancer* 2006;14(5):427-35.
36. Ganz PA, Hahn EE. Implementing a survivorship care plan for patients with breast cancer. *J Clin Oncol* 2008;26(5):759-67.
37. Grunfeld E. Optimizing follow-up after breast cancer treatment. *Curr Opin Obstet Gynecol* 2009;21(1):92-6.
38. Breivik H, Cherny N, Collett B, de Conno F, Filbet M, Foubert AJ, et al. Cancer-related pain: a pan-European survey of prevalence, treatment, and patient attitudes. *Ann Oncol* 2009;20(8):1420-33.
39. Smith T, Stein KD, Mehta CC, Kaw C, Kepner JL, Buskirk T, et al. The rationale, design, and implementation of the American Cancer Society's studies of cancer survivors. *Cancer* 2007;109(1):1-12.
40. van de Poll-Franse LV, Horevoorts N, van Eenbergen M, Denollet J, Roukema JA, Aaronson NK, et al. The Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry: scope, rationale and design of an infrastructure for the study of physical and psychosocial outcomes in cancer survivorship cohorts. *Eur J Cancer* 2011;47(14):2188-94.
41. http://www.e-cancer.fr/en/cancer-plan-2009-2013, 2013.
42. Marino P, Luis Sagaon T, Laetitia M, Anne-Gaelle le CS. Sex differences in the return-to-work process of cancer survivors 2 years after diagnosis: results from a large French population-based sample. *J Clin Oncol* 2013;31(10):1277-84.
43. Mak AK, Chaidaroon S, Fan G, Thalib F. Unintended consequences: the social context of cancer survivors and work. *J Cancer Surviv* 2014.
44. Joutard X, Paraponaris A, Sagaon-Teyssier L, Ventelou B. Continuous-Time Markov Model for Transitions Between Employment and Non-Employment: The Impact of a Cancer Diagnosis. *Annals of Economics and Statistics* 2012;107/108.
45. de Boer AG, Taskila T, Ojajarvi A, van Dijk FJ, Verbeek JH. Cancer survivors and unemployment: a meta-analysis and meta-regression. *Jama* 2009;301(7):753-62.
46. Le Corroller-Soriano AG, Bouhnik AD, Preau M, Malavolti L, Julian-Reynier C, Auquier P, et al. Does cancer survivors' health-related quality of life depend on cancer type? Findings from a large French national sample 2 years after cancer diagnosis. *Eur J Cancer Care (Engl)* 2011;20(1):132-40.
47. Le Corroller-Soriano A, Malavolti L, Mermillod C. *La vie deux ans après le diagnostic du cancer: une enquête en 2004 sur les conditions de vie des malades*. Paris: La documentation française, 2008.

48. Mancini J, Rey D, Preau M, Le Corroller-Soriano AG, Moatti JP. Barriers to procreational intentions among cancer survivors 2 years after diagnosis: a French national cross-sectional survey. *Psychooncology* 2011;20(1):12-8.
49. Preau M, Bouhnik AD, Rey D, Mancini J. Two years after cancer diagnosis, which couples become closer? *Eur J Cancer Care (Engl)* 2011;20(3):380-8.
50. Protiere C, Moumjid N, Bouhnik AD, Le Corroller Soriano AG, Moatti JP. Heterogeneity of cancer patient information-seeking behaviors. *Med Decis Making* 2012;32(2):362-75.
51. Hewitt M, Greenfield S, Stovall E. *From Cancer Patient to Cancer Survivor: Lost in Transition*. Washington, DC: The National Academies Press, 2005.
52. Moran JR, Short PF, Hollenbeak CS. Long-term employment effects of surviving cancer. *J Health Econ*;30(3):505-14.
53. Belot A, Grosclaude P, Bossard N, Jougla E, Benhamou E, Delafosse P, et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2008;56(3):159-75.
54. Groves RM, Biemer PP, Lyberg LE, Massey JT, Nicholls WL, Waksberg J. *Telephone Survey Methodology*: John Wiley & Sons, 2001.
55. Beck F, Gautier A. *Baromètre cancer 2010*. Saint-Denis: INPES, 2012.
56. https://lesdonnees.e-cancer.fr/les-fiches-de-synthese/29-incidence-mortalite/38-ensemble-des-cancers/22-epidemiologie-des-cancers-en-france-metropolitaine-incidence-et-mortalite.html#ind2, 2013.
57. Mazeau-Woynar V, Cerf N. *Survie attendue des patients atteints de cancers en France: état des lieux*. Paris: Institut national du cancer, 2010.
58. Jourdan-Da Silva N. *Instauration d'une insulinothérapie dans le diabète de type 2: données épidémiologiques, étude entred 2001-2003*, 2005.
59. Burdine JN, Felix MR, Abel AL, Wiltraut CJ, Musselman YJ. The SF-12 as a population health measure: an exploratory examination of potential for application. *Health Serv Res* 2000;35(4):885-904.
60. Groenvold M, Klee MC, Sprangers MA, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. *J Clin Epidemiol* 1997;50(4):441-50.
61. Portenoy R. Development and testing of a neuropathic pain screening questionnaire: ID Pain. *Curr Med Res Opin* 2006;22(8):1555-65.
62. Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, et al. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain* 2005;114(1-2):29-36.
63. Lenormand F. Le système d'information de l'assurance maladie, le SNIIRAM et les échantillons de bénéficiaires. . *Courrier des Statistiques* 2005;113-114:33-51.
64. Carstairs V, Morris R. *Deprivation and Health in Scotland*. Aberdeen: Aberdeen University Press, 1991.
65. Rey G, Rican S, Jougla E. Mesure des inégalités de mortalité par cause de décès. Approche écologique à l'aide d'un indice de désavantage social. *Bulletin Epidémiolgique Hebdomadaire* 2011;8-9.
66. http://www.e-cancer.fr/component/docman/doc_download/12164-la-vie-deux-ans-apres-un-diagnostic-de-cancer-de-lannonce-a-lapres-cancer-2014-annexes, 2014.
67. The American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition: AAPOR, 2011.

68. Dolbeault S, Bredart A. Cancer Diagnosis Disclosure: The French Experience. *New Challenges in Communication with Cancer Patients*. New-York: Springer, 2013.
69. Stanton AL. Psychosocial concerns and interventions for cancer survivors. *J Clin Oncol* 2006;24(32):5132-7.
70. Blotiere PO, Weill A, Ricordeau P, Alla F, Allemand H. Perforations and haemorrhages after colonoscopy in 2010: a study based on comprehensive French health insurance data (SNIIRAM). *Clin Res Hepatol Gastroenterol* 2014;38(1):112-7.
71. Fagot JP, Boutrelle A, Ricordeau P, Weill A, Allemand H. HPV vaccination in France: uptake, costs and issues for the National Health Insurance. *Vaccine* 2011;29(19):3610-6.
72. Neumann A, Maura G, Weill A, Ricordeau P, Alla F, Allemand H. Comparative effectiveness of rosuvastatin versus simvastatin in primary prevention among new users: a cohort study in the French national health insurance database. *Pharmacoepidemiol Drug Saf* 2014;23(3):240-50.
73. Santin G, Geoffroy B, Benezet L, Delezire P, Chatelot J, Sitta R, et al. In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse. *J Clin Epidemiol* 2014;67(6):722-30.
74. Elliott J, Fallows A, Staetsky L, Smith PW, Foster CL, Maher EJ, et al. The health and well-being of cancer survivors in the UK: findings from a population-based survey. *Br J Cancer* 2011;105 Suppl 1:S11-20.
75. Friedenreich CM, Howe GR, Miller AB. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. *Ann Epidemiol* 1991;1(5):439-53.
76. Friedenreich CM, Howe GR, Miller AB. The effect of recall bias on the association of calorie-providing nutrients and breast cancer. *Epidemiology* 1991;2(6):424-9.
77. Giovannucci E, Stampfer MJ, Colditz GA, Manson JE, Rosner BA, Longnecker MP, et al. Recall and selection bias in reporting past alcohol consumption among breast cancer cases. *Cancer Causes Control* 1993;4(5):441-8.
78. Vrijheid M, Deltour I, Krewski D, Sanchez M, Cardis E. The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. *J Expo Sci Environ Epidemiol* 2006;16(4):371-84.

Table 1. Sampling design: age at diagnosis & tumour sites strata (VICAN).

Age at diagnosis:	18-52 years old			53-82 years old		
Cancer type:	% in ALD file	expected % in sample	expected size	% in ALD file	expected % in sample	expected size
Breast cancer	40.6%	26.7%	800	17.1%	16.7%	500
Prostate cancer	___	___	___	39.2%	20.0%	600
Melanoma	7.3%	10.0%	300	2.1%	5.0%	150
Thyroid cancer	9.7%	10.0%	300	___	___	___
Colorectal cancer	8.7%	11.7%	350	14.2%	10.0%	300
Upper aerodigestive tract cancers	9.8%	10.0%	300	4.6%	6.7%	200
Bladder cancer	___	___	___	3.7%	6.7%	200
Kidney cancer	3.2%	5.0%	150	3.1%	5.0%	150
Non-Hodgkin lymphoma	5.3%	6.7%	200	3.5%	6.7%	200
Cervical cancer	4.3%	6.7%	200	2.5%	6.7%	200
Endometrial cancer	___	___	___	0.5%	3.3%	100
Lung cancer	11.1%	13.3%	400	9.5%	13.3%	400
Total:	100%	100%	3000	100%	100%	3000

Reading example: among patients aged 18-52 at diagnosis, for the 9 selected tumour sites, 40.6% of patients registered in the ALD file in 2012 had breast cancer. The expected figure was much lower in the sample: 26.7% (N=800).

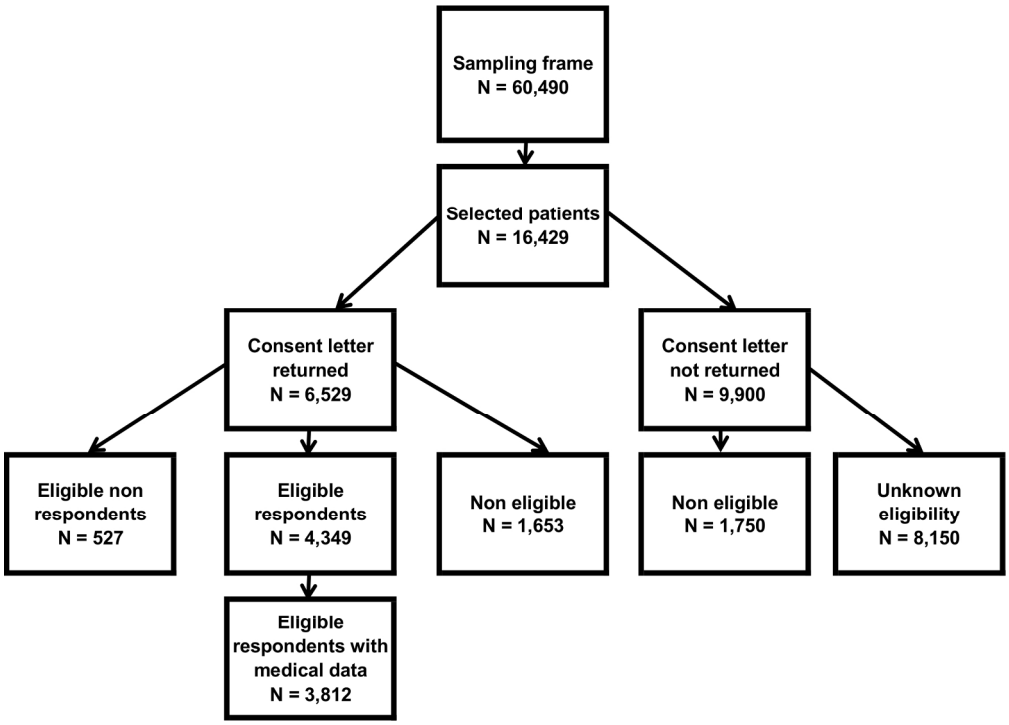
Table 2. Final sample: age at diagnosis & tumour sites strata (VICAN).

	Age at diagnosis:		
Cancer type:	18-52	53-82	*Total*
Breast cancer	971	379	1,350
Prostate cancer	___	479	479
Melanoma	162	114	276
Thyroid cancer	181	___	181
Colorectal cancer	258	229	487
Upper aerodigestive tract cancers	153	131	284
Bladder cancer	___	143	143
Kidney cancer	108	110	218
Non-Hodgkin lymphoma	163	122	285
Cervical cancer	97	78	175
Endometrial cancer	___	75	75
Lung cancer	136	260	396
Total:	2,241	2,108	4,349

Table 3. Comparison between eligible respondents and eligible non-respondents (VICAN).

	respondents (N=4,349)	non-respondents (N=527)
	% column	
Gender:		
-men	47.2%	58.9%
-women	52.8%	41.1%***
Age: mean (SE)	60.4 (11.4)	64.7 (11.5)***
Social Deprivation Index:		
-<first quartile	20.3%	14.4%
- [1st-3rd quartiles]	33.4%	29.1%
->third quartile	46.3%	56.5%***
Cancer type:		
-breast cancer	35.1%	22.4%
-prostate cancer	24.8%	29.6%
-melanoma	3.1%	3.1%
-thyroid cancer	2.0%	1.4%
-colorectal cancer	11.8%	13.3%
-upper aerodigestive tract cancers	4.0%	9.1%
-bladder cancer	4.5%	4.8%
-kidney cancer	3.4%	3.6%
-non-Hodgkin lymphoma	3.2%	2.9%
-cervical cancer	1.3%	0.9%
-endometrial cancer	1.8%	0.8%
-lung cancer	4.9%	8.1%***
Cancer progression since diagnosis:	82.6%	78.2%
-no	17.4%	21.8%***
-yes		

***, **, *, ns: respectively statistically significant at p<0.001, p<0.01, p<0.05, not significant (Student's t-test for age, Pearson's χ^2 for other variables).



Sample selection pathway (VICAN).
187x134mm (300 x 300 DPI)

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	✓ 1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
	✓	(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	✓ 2	Explain the scientific background and rationale for the investigation being reported
Objectives	✓ 3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	✓ 4	Present key elements of study design early in the paper
Setting	✓ 5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	✓ 6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants
	not applicable	(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	✓ 7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
	not applicable .	
Bias	✓ 9	Describe any efforts to address potential sources of bias
Study size	✓ 10	Explain how the study size was arrived at
Quantitative variables	✓ 11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	✓ 12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy
		(*e*) Describe any sensitivity analyses

Continued on next page

Results			
Participants	✓	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
			(b) Give reasons for non-participation at each stage
			(c) Consider use of a flow diagram
Descriptive data	✓	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
			(b) Indicate number of participants with missing data for each variable of interest
			(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)
Outcome data		15*	*Cohort study*—Report numbers of outcome events or summary measures over time
not applicable.			*Case-control study*—Report numbers in each exposure category, or summary measures of exposure
			Cross-sectional study—Report numbers of outcome events or summary measures
Main results		16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
not applicable.			(*b*) Report category boundaries when continuous variables were categorized
			(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses		17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
not applicable.			
Discussion			
Key results	✓	18	Summarise key results with reference to study objectives
Limitations	✓	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	✓	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
not applicable.			Generalisability 21 Discuss the generalisability (external validity) of the study results
Other information			
Funding	✓	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.