HAL open science

Completeness of a newly implemented general cancer registry in northern France: Application of a three-source capture-recapture method

S.D. Plouvier, P. Bernillon, K. Ligier, D. Theis, P.-H. Miquel, D. Pasquier, L.-P. Rivest

To cite this version:

S.D. Plouvier, P. Bernillon, K. Ligier, D. Theis, P.-H. Miquel, et al.. Completeness of a newly implemented general cancer registry in northern France: Application of a three-source capture-recapture method. Epidemiology and Public Health = Revue d'Epidémiologie et de Santé Publique, 2019, 67, pp.239 - 245. 10.1016/j.respe.2019.04.053 . hal-03488179

HAL Id: hal-03488179

https://hal.science/hal-03488179v1

Submitted on 20 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Completeness of a newly implemented general cancer registry in northern France: application of a three-source capture-recapture method

Exhaustivité d’un registre général de cancers récemment implanté dans le Nord de la France: application d'une méthode capture-recapture à trois sources

TITRE COURT: exhaustivité registre

Sandrine D Plouvier* a

Pascale Bernillon b

Karine Ligier a

Didier Theis c

Pierre-Henri Miquel d

David Pasquier e,f

Louis-Paul Rivest g

a Registre général des cancers de Lille et de sa région, C2RC, Lille, France

e available under the CC BY NC user license

b Santé Publique France, Saint-Maurice, France

c Département d'information médicale, Administration générale, CHR&U, Lille, France

d Ramsay générale de santé, Hôpital privé La Louvière, Lille, France

e Academic Radiation Oncology Department, Centre Oscar Lambret, Lille University, Lille, France

f CRISTAL, UMR CNRS 9189, Lille, France

g Department of mathematics and statistics, Université Laval, Quebec, Quebec, Canada

Correspondance

Sandrine D Plouvier* Registre général des cancers de Lille et de sa région Hôpital Calmette Bld du Professeur J Leclercq 59000 Lille France

splouvier@registrecancers59.fr

Pascale Bernillon Agence nationale de santé publique 12 rue du Val d'Osne 94415 Saint-Maurice France

Pascale.BERNILLON@santepubliquefrance.fr

Karine Ligier Registre général des cancers de Lille et de sa région Hôpital Calmette Bld du Professeur J Leclercq 59000 Lille France

KLIGIER@registrecancers59.fr

Didier Theis Département d'Information Médicale CHRU Lille 2 avenue Oscar Lambret 59000 Lille France

Didier.THEIS@CHRU-LILLE.FR

Pierre-Henri Miquel Hopital Privé La Louvière 69 rue de la Louvière 59000 Lille

p.miquel@ramsaygds.fr

David Pasquier Service de radiothérapie Centre Oscar Lambret 3 rue Frédéric Combemale 59000 Lille France

d-pasquier@o-lambret.fr

Louis-Paul Rivest Université de Laval 2325 rue de l'Université Québec Canada

Louis-Paul.Rivest@mat.ulaval.ca

Completeness of a newly implemented general cancer registry in northern France: application of a three-source capture-recapture method

Exhaustivité d'un registre général de cancers récemment implanté dans le Nord de la France: application d'une méthode capture-recapture à trois sources

Abstract

Background: Completeness, timeliness and accuracy are important qualities for registries. The objective was to estimate the completeness of the first two years of full registration (2008/2009) of a new population-based general cancer registry, at the time of national data centralisation.

Methods: Records followed international standards. Numbers of cases missed were estimated from a three-source (pathology labs, healthcare centres, health insurance services) capture-recapture method, using log-linear models for each gender. Age and place of residence were considered as potential variables of heterogeneous catchability.

Results: When data were centralized (2011/2012), 4446 cases in men and 3642 in women were recorded for 2008/2009 in the Registry. Overall completeness was estimated at 95.7% (95% CI: 94.3-97.2) for cases in men and 94.8% (95% CI: 92.6-97.0) in women. Completeness appeared higher for younger than for older subjects, with a significant difference of 4.1% (95% CI: 1.4-6.7) for men younger than 65 compared with their older counterparts. Estimates were collated with the number of cases registered in 2014 for the years 2008/2009 (4566 cases for men/3755 for women), when additional structures had notified cases

retrospectively to the Registry. These numbers were consistent with the stratified capture-recapture estimates.

Conclusion: This method appeared useful to estimate the completeness quantitatively. Despite a rather good completeness for the new Registry, the search for cases among older subjects must be improved.

Key words: Capture-Recapture. Completeness. Registry. Age

Résumé:

Position du problème: L'exhaustivité, la rapidité et la précision de l'enregistrement sont des qualités primordiales pour les registres. L'objectif de l'étude était d'estimer l'exhaustivité des premières années d'enregistrement (2008/2009) d'un registre général de cancers, en population générale, à l'heure de la centralisation nationale des données de registres des cancers.

Méthodes: Les données étaient codées et enregistrées selon les recommandations internationales et nationales. Le nombre de cas manquants a été estimé à partir d'une méthode capture-recapture à trois sources (laboratoires d'anatomocytopathologie, établissements de santé, services d'assurance-maladie), à l'aide de modèles log-linéaires. Les hommes et les femmes ont été étudiés séparément. L'âge et le lieu de résidence ont été considérés comme potentielles variables d'hétérogénéité de capture.

Résultats: Lors de la centralisation nationale des données (2011/2012), 4446 cas étaient enregistrés chez les hommes et 3642 chez les femmes pour 2008/2009. L'exhaustivité globale du registre était estimée à 95,7% (IC 95% : 94,3-97,2) chez les hommes et 94,8% (IC 95% : 92,6-97,0) chez les femmes. L'exhaustivité était meilleure chez les sujets les plus jeunes, avec une différence significative de 4,1% (IC 95% : 1,4-6,7) pour les hommes de moins de 65 ans comparés aux plus

âgés. Le nombre de cas estimé a été comparé au nombre de cas enregistrés quelques années plus tard, en 2014, pour 2008/2009 (4566 cas pour les hommes / 3755 pour les femmes). Des structures supplémentaires avaient alors notifié rétrospectivement au Registre des cas pour les années initiales. Les enregistrements concordaient avec les estimations stratifiées obtenues par la méthode capture-recapture.

Conclusion: Cette méthode est apparue utile pour quantifier l'exhaustivité du Registre. Cette exhaustivité lors de la centralisation nationale semblait assez bonne pour un registre nouvellement implanté. La recherche de cas incidents parmi les sujets plus âgés devait néanmoins être améliorée.

Mots-Clés: Capture-Recapture. Exhaustivité. Registre. Âge

Introduction

The data from cancer registries are extensively used for public health policy development (care planning), monitoring (incidence, survival, patient management, cluster detection) and research purposes (observational and etiological studies). At the forefront of the requested qualities for registration are completeness - the proportion of incident cases registered among all incident cases -, accuracy and timeliness[1,2] .

In France, the *Registre général des cancers de Lille et de sa region* (general cancer registry of Lille and its area – RLA[3]) fell within the framework of the creation of new cancer registries in urban areas following the first National Cancer Plan[4], to improve the representativeness of French registries. The RLA was designed to collect cancer cases occurring from January 1, 2005 onwards. Registration, at first restricted to solid tumours in people aged 15 or more, has covered all cases occurring since January 1, 2008 regardless of age and type.

In this context of a newly implemented registry, the objective of the study was to assess the completeness of the first two years (2008 and 2009) of full registration of invasive cancers in the RLA, at the time of national data centralisation (February, 2011 and February, 2012 for 2008 and 2009 respectively), using a three-source capture-recapture (CR) method. A secondary objective was to discuss the estimated completeness in the light of cases registered in 2014 for these years, when additional entities provided information on cancer cases retrospectively to the RLA for 2008 and 2009.

Methods

1-The Registry

1 1-Implementation and notification of cancer cases to the registry

The RLA is a population-based general cancer registry covering a geographical area in Northern France of approximately 780 000 inhabitants, the “Lille area’’ (LA). This dense urban area is a health territory of 101 towns. It also corresponds to a territorial division of the national statistics office, Insee (*Institut National de la Statistique et des Études Économiques)*[5]. Detailed annual population data for the LA is provided by this office using the Omphale method.

The Registry has the agreement of the French national ethical committee (*Commission Nationale de l’Informatique et des Libertés*) and the national committee of evaluation of registries (*Comité d’Evaluation des Registres)*. Every subject receive information concerning the registration, and can refuse to be recorded.

When cases incident in 2008 and 2009 were investigated, several structures independently notified cancer cases to the RLA . All private and public hospitals located in the LA, plus some of the main ones outside the coverage zone, notified hospitalisations of patients suffering from cancer. All radiotherapy centres located in the LA, plus those northeast of the area, notified treatments for malignant tumours. The main health insurance services, covering about 90% of the regional population, transmitted cases of exemption from prescription charges related to cancer. Three out of five pathology labs located in the LA, plus one located to the northeast provided detailed reports of analyses concerning malignant tumours. This notification scheme implied a higher risk of being missed by the registry for subjects for whom diagnosis and treatment were made without hospitalisation. To minimise this, an active search for cases eligible for registration was done in

specific structures (some dermatologic oncologic multidisciplinary team meetings; some haematology labs; some medical files for outpatients in health care centres). It led to the identification of unreported incident cases.

In the following years, the last two pathology labs out of five located in the LA, the main hospitals and two radiotherapy centres located around the LA enrolled in the notification process. The data provided retrospectively to the RLA from 2005 onwards were investigated.

Each notification provided dates, International Classification of Disease-tenth edition (ICD-10) codes or the specific pathological code of the affection, and full names, date of birth, sex and address of the subject.

1 2-Search for duplicates

The matching process was performed at the individual level in the database. Search for duplicates included cross-referencing of names/forename/sex/date of birth. For ambiguous cases, manual authentications were requested.

1 3-Authentication and registration of incident cases

Each specific disease code and pathological report was investigated. Medical files were inspected to confirm the registration of a case.

All cases were coded and registered according to the rules of the French network of cancer registries (Francim), the European Network of Cancer Registries (ENCR), and the International Association of Cancer Registries (IACR)[6,7]. For each case, the date of diagnosis, topography, morphology and basis of diagnosis were recorded as well as the name, forename, date of birth, gender, and address at diagnosis of the patient. Specific software, DEPEdits/IARC Tools, developed by

the International Agency for Research on Cancer (Iarc) was used to check the consistency of coding.

2-Cancer cases

Incident invasive cancers occurring between January 1, 2008 and December 31, 2009 in the LA and registered at the time of national data centralisation were considered in the study. In France, the data of population-based registries are centralized for any year N in February, year N+3, in order to estimate national incidence rates[8]. Cases incident in 2008 are centralized in February 2011 and those incident in 2009 are centralized in February 2012 in order to produce these national estimates. So, cases recorded in February 2011 and in February 2012 were considered respectively for 2008 and for 2009.

3-Sources for the capture-recapture study

For cases incident in 2008, notifications received between December 1, 2007 and December 31, 2009 were considered in the CR study. For cases incident in 2009, it was those received between December 1, 2008 and December 31, 2010.

Three sources were considered for each case in the analyses:

- health care centres (HCC) : coded 1 if at least one private and/or one public hospital and/or one radiotherapy centre had notified the RLA of this cancer, 0 otherwise,
- pathology labs (L) : coded 1 if at least one pathology lab had notified the RLA of this cancer, 0 otherwise,
- health insurance services (HI): coded 1 if at least one health insurance service had notified the RLA of this cancer, 0 otherwise.

4-Possible variables of heterogeneous catchability

Age and place of residence at diagnosis were investigated as potential variables of heterogeneous catchability. The hypothesis for age was that young patients of the LA would have a higher probability of being notified than older patients. Men and women had different cancer profiles; thus, distribution of age varied according to gender (first quartile, median and third quartile being 57, 65, 75 and 52, 64 and 78 respectively among men and women). Age distribution was broader among women and they were rather often younger than 50 years old, which was different from cases among men. Thus, cut-off values were defined as at least 65 years old versus less for men, and 52 versus less for women. Concerning place of residence, the hypothesis was that those living at the western edge of the LA were at higher risk of being treated in places that did not yet notify the RLA of cancer cases. Therefore, people who lived at the western edge of the area were studied versus those living in the other parts.

5-Analyses

Men and women were studied separately.

As a preliminary step, a global analysis without covariates[9-11] was performed, assuming that the probability of capture depended only on the sources of ascertainment. Then, analyses incorporating age and place of residence of subjects as possible sources of heterogeneity of capture were undertaken. Estimates from the latter were used to calculate stratified and overall completeness. Log-linear analyses (Poisson regression) were performed using STATA® 11.0 software (College station, Texas)[12].

5 1-Global analysis

First, two-way dependencies between the sources were investigated by calculation of odds ratios (OR) on 2x2 contingency tables according to Wittes' method[13]. Then, the eight feasible log-linear models, taking into account the three sources and possible two-way dependence between these sources (figured as interaction terms), were fitted to the whole set of observed data. Model selection was based on the goodness-of-fit deviance test and Bayesian Information Criterion (BIC).

5 2-Heterogeneous catchability

In a second step, a new model was built to determine whether age was a variable of heterogeneous catchability. Age and all interaction terms between each source and age were added to the full model containing the three sources and all possible two-way dependencies between these sources. Model selection was based on the goodness-of-fit deviance test, Bayesian Information Criterion (BIC) and a combination of significance and epidemiological relevance of interaction terms. A model with a lower BIC could be not kept if it did not contain a relevant interaction term. The same was done with the place of residence instead of age to determine whether residence was a variable of heterogeneous catchability. These models gave an estimate of missing cases and its 95% confidence interval in each relevant stratum and for each gender.

5 3-Completeness

Overall and stratified estimates of missing cases obtained with previous models were established from notified cases; thus, cases recorded without notification

should be accounted for by these estimates. Overall completeness of registration for each sex was calculated as follows:

$$\frac{\text{Total number of registered cases (including those notified and those without notification)}}{\text{Overall stratified estimate of missing cases from the stratified model+Number of notified registered cases}}$$

and stratified completeness was calculated for each stratum as follows:

$$\frac{\text{Number of registered cases in the stratum (including those notified and those not notified)}}{\text{Stratified estimate of missing cases from the model for the stratum+Number of notified registered cases in the stratum}}$$

Their variance and 95% confidence interval were calculated as detailed in the additional material.

Difference in the completeness between two strata for the same sex was evaluated as presented in the additional material using a z statistic.

5 4-2014 Assessment

The consistency of the results of the CR study performed was discussed in the light of incident cases registered for the years 2008 and 2009 in the RLA in February 2014.

Results

1-Incident cancer cases

At the time of national data centralisation (2011 and 2012), 4 446 incident cancers were registered for men and 3 642 for women for 2008 and 2009; 57 of these

cases among men (1.3%, mainly haematological, prostate and skin cancers) and 35 of these cases among women (1%, mainly haematological and skin cancers) had not been notified. The distribution of age and place of residence of the notified and non-notified cases according to gender are presented in Table 1. The distribution of notifications by gender is shown in Figure 1. The HCC source notified almost all cases (91.8 and 93.9% of cases among men and women respectively) whereas the L source notified the fewest cases (48.7 and 51.8% of cases among men and women respectively).

2-Global analysis

Calculating odds ratios (OR) by Wittes' method showed a positive dependence between the two sources HCC and L for men and women (OR 1.47 (95%CI: 1.10-1.96) and 2.72 (95%CI: 1.84-4.02) respectively) and also between HCC and HI for men and women (OR 3.65 (95%CI: 2.73-4.89) and 6.22 (95%CI: 4.13-9.37) respectively). Both L and HI were more likely to notify cases notified by HCC than cases not notified by HCC. No dependence between HI and L could be identified (OR 1.10 (95%CI: 0.96-1.25) among men and 1.00 (95%CI: 0.86-1.16) among women).

Table 2 presents properties (number of degrees of freedom, deviance, Bayesian Information Criterion, p-value associated with the goodness-of-fit test) and overall estimates of missing cases with their 95% confidence intervals given by the eight feasible log-linear models for each gender, calculated from notified cases. The model that best fitted the data according to the chosen criteria was that in which the above-mentioned dependence results were taken into account. It gave an

estimate of 237.0 (95%CI: 172.0-326.8) missing cases among men and 226.0 (95%CI: 146.5-348.6) among women for both 2008 and 2009.

3-Variables of heterogeneous catchability

Probabilities of capture of a case varied with age, but not with place of residence (details not shown). All interaction terms identified in the global analysis were still significant in this analysis with age. For men, the model with the lower BIC was that in which all interactions terms between age and each source were significant as well as the dependence between HI and HCC (p=0.17, BIC=-34.14), the next model was that that comprised in addition a significant dependence between the sources HCC and L (p=0.61, BIC=-30.84). Thus, the latter was chosen for the pertinence of this dependence. For women, the model with the lower BIC included the significant dependence between the source HCC and HI and between HCC and L and significant interaction terms between age and L and between age and HI. Table 3 presents models, estimates of missing cases and their 95% confidence intervals according to sex and age, calculated from notified cases. For men, the probability of being notified diminished for the older age group, whatever the source. The stratified estimates of missing cases were 63.8 (95%CI: 44.3-91.9) for those under 65 years old and 191.6 (95%CI: 137.9-266.1) for those 65 or older. For women, the probability of being notified by sources L and HI diminished for the older age group. The stratified estimates of missing cases in women were 33.4 (95%CI: 21.1-52.7) for those under 52 and 201.9 (95%CI: 130.6-312.1) for those 52 or older.

5-Completeness

Overall completeness was estimated at 95.7 (95%CI: 94.3-97.2) for men and 94.8 (95%CI: 92.6-97.0) for women from the stratified analysis. Table 3 presents overall and age-group stratified completeness. Completeness of the RLA appeared higher for the younger than for the older people for each sex; the difference was significant among men: 4.1% (95%CI: 1.4-6.7) but not among women: 2.4% (95%CI: -1.0-5.8).

6-Assessment in 2014

In February 2014, 4566 incident cancers had been recorded for men and 3755 for women for the 2008-2009 study period. Compared to the records at the time of national data centralisation (2011 and 2012), this corresponded to a global increase of 2.7% for cases among men (38 cases (+1.8%) under 65 years old and 82 (+3.4%) for those older) and 3.1% for cases among women (20 cases (+2.3%) under 52 years old and 93 (+3.2%) for those older). Most of the additional cases had been provided by entities that joined the notification process late and were related to cases treated in 2008 and 2009 by these entities. Few of the additional cases had been notified later than two years after diagnosis because of either a diagnosis made outside a clinical centre or an unwillingness of the person to be treated. None of the number of cases by stratum was higher than the capture-recapture estimates (which was 63.8 for men under 65 years old and 191.6 for those older and 33.4 for women under 52 years old and 201.9 for those older). Additional cases recorded in men under 65 years old were mainly melanoma of the skin (n=12), prostate (n=6) and upper aerodigestive tract cancers (n=5); in older men, they were mainly squamous cell carcinoma of the skin (n=25), prostate (n=14) and lung (n=8) cancers. In women under 52 years old, they were mainly

melanoma of the skin (n=11) and breast cancers (n=8); in older women, they were mainly squamous cell carcinoma of the skin (n=27), breast cancers (n=21), non-Hodgkin lymphoma (n=10), melanoma of the skin (n=8) and colorectal cancers (n=6).

Discussion

The completeness at the time of national data centralisation (2011 and 2012) of the first two years registered, based on estimates from the three-source CR study appeared globally rather good, but lower in the strata of older subjects. The number of cases registered for these two years after additional years had elapsed did not contradict these estimates.

Completeness, with accuracy and timeliness, are important for a registry. Some instances edited rates of completeness to be reached within a given timeframe[1]. In European registries affiliated to the ENCR, this evaluation is still registry-dependent in practice[14]. Among 116 of these registries, periodicity of and methods used for evaluations varied widely; for 53%, it involved at least the use of one quantitative method (giving rise to a numerical evaluation of the completeness).

However, the Iarc referenced several methods for assessing completeness[15]. Methods rely on data sources (mean number of notifications per case, proportion of cases microscopically verified, use of death certificates (DC) - for estimating the number of missed cases, for example -), or on independent case ascertainment (comparisons with independent files, case-finding studies, CR methods, and mortality/incidence ratio per site) or on historic data (based on time trends observed). More recently, the flow method gave an estimation of the number of

missed cases, in both surviving and dead subjects, using the probability of survival, the probability of the cancer being mentioned on the DC and the probability of non recording of the case[16].

Estimations of selected-site or overall completeness of cancer registries based on quantitative and/or qualitative methods have been published[3, 17-25]. Each method has its strengths and limitations. Combining them would certainly be a good practice. In France, nominative DC cannot be made available mandatorily for cancer registries. In a newly created registry, historic data methods cannot be exploited. Thus, CR method appeared suitable to achieve the objective of the present study, with the advantage of quantifying the completeness. In addition, such methods can rely on sources of information routinely employed by the registry, thus they do not require additional procedure[15].

The assumptions and limits of CR methods have been described[9-11, 26-29]. All cases here were verified: they were true cases and arose in the area during the period studied. As the sources of notification are involved at different times in caring for the subject, dates of notifications vary between sources. Yet, incidence date recorded is usually the date of the pathological proof if any. Health insurances services can exempt from prescription charges as soon as a person has a diagnostic exam. This date of exam, possibly before the pathological proof, correspond to the date of notification to the registry. On the contrary, hospital stay often occur after this proof, which can be obtained on sample made in ambulatory medicine. This is especially important for cases incident very early or very late in a year N, for which some notification can occur the previous year (N-1) or the next year (N+1). This led us to consider an extended period for notifications to ensure an accurate distribution of sources for a case. The coupled computed and manual

matching procedures should have minimized the number of both missed duplicates and mismatches, even though there was no common identifier across the sources. Although the population studied was not a closed population, it should not have had noticeable consequences on the result. Dependence between sources (the probability of being notified by a source varying with the probability of being notified by another one), likely in a cancer registry, could lead to under/overestimation of missing cases, and thus over/underestimation of the completeness, depending on the type of dependence. This is also the case for variables of heterogeneous catchability (characteristics of a case that modify the probability of the case being notified by a given source). As three sources were available, dependence between the sources, as well as some variables of heterogeneous catchability could be taken into account in log-linear models. Nonetheless, other characteristics, which are not routinely assessed in the RLA, such as extent of disease at diagnosis or comorbidities, could also have been variables of heterogeneous catchability. It is noteworthy that the cancer site can also play such a role. Some lesions can be treated only in outpatient care, some could be not related to the health assurance services, restricting the possibility of being notified to the registry. The various cancer sites and the small number of persons for some of them preclude the ability to study this influence in a sole model. Therefore, the influence of cancer site on the estimates is not quantified here. Nonetheless, estimating completeness by cancer site represents a prospect following this first CR study.

In the final models, the positive dependence between the HCC and L sources, and between HCC and HI, was consistent with the notifications schemes at the implementation of the RLA. It is not unlikely that, with the participation of additional

pathology centres, a positive dependence could also be seen between HI and L. The number of cancers incident in 2008 and 2009, as assessed in 2014, did not contradict the estimates of missing cases from the CR study performed.

The overall completeness of the RLA appeared to be in accordance with those reported for other registries for the same period. Based on the flow method, the overall completeness of the Austrian National Cancer Registry was 94.2% for 2005 after five years of registration, and 92.1% for 2006 after three years of registration[20]. Using a two-source (clinician and pathologist/DC) CR method, the overall completeness of the Bulgarian National Cancer Registry was 94.7% for the period 2006 to 2010, as estimated in 2012[19]; in addition, completeness was lower for women than for men. With independent case ascertainment (records from hospital discharge received in 2005), the overall completeness of the Icelandic Cancer Registry was found to be somewhat better at 99.2% for the period 2000-2001[24]. Also, at the Cancer Registry of Norway, the overall completeness reached 98.8% for the period 2001-2005 based on a two-source (among pathology, DC, Clinic) CR method and 97.8% for 1999 after five years using the flow method [21].

In the study presented completeness was lower in the strata for older subjects, as expected, mainly for men. Invasive investigation, heavy treatment can be avoided in older, weak people. Reporting to HI services could also be less common among them. Various variations in completeness with age have been reported previously[17, 22,30]. The number of additional cases obtained with retrospective data from additional entities was greater for men and women in the group “older people”. Nonetheless, as estimates of missing cases were greater among this age group than among younger subjects, in proportion, fewer people were retrieved

among the older group than among the younger one. Attention should therefore be paid to cases among “older” people. If specific entities such as oncology multidisciplinary team meetings, ultrasound/scanner centres or general practitioners could join the notification process, the registration would probably improve. People not notified were probably people advanced in age, or with comorbidities that would impair treatment, or who died soon after diagnosis, for whom investigations were limited. Few oncology multidisciplinary team meetings have records that permit the selection of people eligible for registration. These will be approached to join in the notification entities. The others, as well as ultrasound/scanner centres and general practitioners, not having such records, cannot routinely be entities of notification; a manual search of eligible cases would be too costly in time. Nonetheless, when data from the Swedish cancer register were crossed with those from 2009 in the Swedish register of palliative care, only a small minority of cases unknown by the cancer registry came from older patients cared for by general practitioners in nursing homes[30]. Most were related to cancers affecting organs easily visualized by imaging but hard to reach for samples.

In the study presented, no variation in completeness with place of residence was detected, contrary to our hypothesis.

Finally, in registries, variations in completeness with time[19,20,24] and site[19-21,24] are well known. Additional entities conveyed substantial information to the RLA: whilst allowing retrieval of those cases of expected types - tumours that do not usually require hospitalisation like skin cancers, or that were managed without it, especially among old or weak patients as in some breast and prostate cancers, - they also enabled retrieval of cases of various profiles cared for outside the LA.

In conclusion, three-source CR method appeared useful for estimating the completeness rate of the registry quantitatively. An upcoming analysis by cancer site will give more detailed completeness figures. Despite a new implementation, the completeness of the RLA appeared rather good. Attention should nonetheless be paid to cases among older people and the search for such cases needs to be improved.

Funding:
This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sector

Conflicts of interest: None.

References

1 Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer 2009;45: 747-55.

2 Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer 2009;45: 756-64.

3 Ligier K, Plouvier S, Danzon A, Martin P, Benoit E, Molinie F, et al. Éléments d'exhaustivité et résultats pour la première année d'enregistrement du Registre general des cancers de Lille et de sa region. Rev Epidemiol Sante Publique 2012;60: 131-9.

4 Mission interministerielle pour la lutte contre le cancer. Plan cancer 2003-2007. 44 p.

5 Statistics.Insee. [cited oct 2018]. Available at: https://www.insee.fr/fr/statistiques?debut=0&theme=1&geo=ZE2010-3111

6 European Network of Cancer Registries. Recommandations and working groups. [cited oct 2018]. Available at: https://www.encr.eu/recommandations-and-working-groups.

7 International Association of Cancer Registries. IACR Standards. [cited oct 2018]. Available at: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=105&Itemid=472.

8 Binder-Foucard F, Bossard N, Delafosse P, Belot A, Woronoff AS, Remontet L, French network of cancer registries. Cancer incidence and mortality in France over the 1980-2012 period: solid tumors. Rev Epidemiol Sante Publique 2014;62: 95-108.

9 Gallay A, Nardone A, Vaillant V, Desenclos JC. The capture-recapture applied to epidemiology: principles, limits and application. Rev Epidemiol Sante Publique 2002;50: 219-32.

10 Rivest LP. Théorie et applications des modèles de capture-recapture. Bulletin AMQ 2013;LIII: 65-78.

11 Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev 1995;17: 243-64.

12 STATA® 11.0 statistical software (College Station, TX)

13 Wittes TJ, Colton T, Sidel SW. Capture-recapture methods for assessing the completeness of cases ascertainment when using multiple information sources Chron Dis 1974;69:93-97.

14 Zanetti R, Schmidtmann I, Sacchetto L, Binder-Foucard F, Bordoni A, Coza D, et al. Completeness and timeliness: Cancer registries could/should improve their performance. Eur J Cancer 2015;51: 1091-8.

15 Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL, Editors. Comparability and quality control in cancer registration. (IARC technical report n°19) ed. Lyon: IARC (WHO) and IACR, 1994.

16 Bullard J, Coleman MP, Robinson D, Lutz JM, Bell J, Peto J. Completeness of cancer registration: a new method for routine use. Br J Cancer 2000;82: 1111-6.

17 Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. Acta Oncol 2009;48: 27-33.

18 Brewster DH, Crichton J, Harvey JC, Dawson G. Completeness of case ascertainment in a Scottish regional cancer registry for the year 1992. Public Health 1997;111: 339-43.

19 Dimitrova N, Parkin DM. Data quality at the Bulgarian National Cancer Registry: An overview of comparability, completeness, validity and timeliness. Cancer Epidemiol 2015;39: 405-13.

20 Hackl M, Waldhoer T. Estimation of completeness of case ascertainment of Austrian cancer incidence data using the flow method. Eur J Public Health 2013;23: 889-93.

21 Larsen IK, Smastuen M, Johannesen TB, Langmark F, Parkin DM, Bray F,et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer 2009;45: 1218-31.

22 McClish D, Penberthy L. Using Medicare data to estimate the number of cases missed by a cancer registry: a 3-source capture-recapture model. Med Care 2004;42: 1111-6.

23 Moller H, Richards S, Hanchett N, Riaz SP, Luchtenborg M, Holmberg L, et al. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. Br J Cancer 2011;105: 170-6.

24 Sigurdardottir LG, Jonasson JG, Stefansdottir S, Jonsdottir A, Olafsdottir GH, Olafsdottir EJ, et al. Data quality at the Icelandic Cancer Registry: comparability, validity, timeliness and completeness. Acta Oncol 2012;51: 880-9.

25 Thoburn KK, German RR, Lewis M, Nichols PJ, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. Cancer 2007;109: 1607-16.

26 Capture-Recapture and Multiple-Record Systems Estimation I: History and Theoretical Development. International Working Group for Disease Monitoring and Forecasting. Am J Epidemiol 1995;142:1047-58.

27 Chang YF, LaPorte RE, Aaron DJ, Songer TJ. The importance of source selection and pilot study in the capture-recapture application. J Clin Epidemiol 1999;52: 927-8; discussion 9-33.

28 Silcocks PB, Robinson D. Completeness of ascertainment by cancer registries: putting bounds on the number of missing cases. J Public Health (Oxf) 2004;26: 161-7.

29 Tilling K. Capture-recapture methods--useful or misleading? Int J Epidemiol 2001;30: 12-4.

30 Nilsson M, Tavelin B, Axelsson B. A study of patients not registered in the Swedish Cancer Register but reported to the Swedish Register of Palliative Care 2009 as deceased due to cancer. Acta Oncol 2014;53: 414-9.

Table 1: Characteristics of the cases registered for 2008 and 2009 at the time of national data centralisation (February 2011 and 2012, respectively)

			Notified	Nor notified
Men	Age	<65 years	2114	19
		≥65 years	2275	38
	Residence	Western	1056	9
		Centre	3333	48
	Total		4389	57
Women	Age	<52 years	853	4
		≥52 years	2754	31
	Residence	Western	811	7
		Centre	2796	28
	Total		3607	35

Table 2: Overall estimates of the number of missing cases for 2008 and 2009 by sex from log-linear models applied to notified cases

Model	df	Deviance	p-value[a]	BIC	Number of missing cases	
					Estimate	(95% CI)
Men						
HI L HCC HI*L HI*HCC L*HCC	0	0	1	0	259.8	(183.7-367.3)
HI L HCC HI*HCC L*HCC	1	0.19	0.17	-6.48	237.0	(172.0-326.8)
HI L HCC HI*L HI*HCC	1	6.90	<0.01	-1.48	177.1	(146.2-214.5)
HI L HCC HI*L L*HCC	1	83.14	<0.0001	74.76	71.1	(58.5- 86.4)
HI L HCC HI*L	2	88.07	<0.0001	71.30	83.2	(73.0- 94.7)
HI L HCC L*HCC	2	83.93	<0.0001	67.15	73.2	(60.9- 88.0)
HI L HCC HI*HCC	2	7.86	0.02	-8.92	169.7	(143.0-201.3)
HI L HCC	3	88.51	<0.0001	63.35	84.6	(75.0-95.4)
Women						
HI L HCC HI*L HI*HCC L*HCC		0	1	0	226.2	(143.0-357.9)
HI L HCC HI*HCC L*HCC		0.00	0.99	-8.19	226.0	(146.5-348.6)
HI L HCC HI*L HI*HCC		28.42	<0.0001	20.23	83.1	(65.0-106.1)
HI L HCC HI*L L*HCC		90.84	<0.0001	82.65	36.4	(28.9- 45.7)
HI L HCC HI*L		91.43	<0.0001	75.04	34.1	(29.0- 40.3)
HI L HCC L*HCC		94.25	<0.0001	77.87	39.2	(31.6- 48.5)
HI L HCC HI*HCC		28.79	<0.0001	12.41	86.0	(69.1-107.0)
HI L HCC		95.09	<0.0001	70.52	36.4	(31.3-42.3)

HI: health insurance services; L: pathology labs; HCC: health care centres; df: number of degrees of freedom; BIC: Bayesian Information Criterion; CI: Confidence interval.

[a] p-value of the goodness-of-fit test.

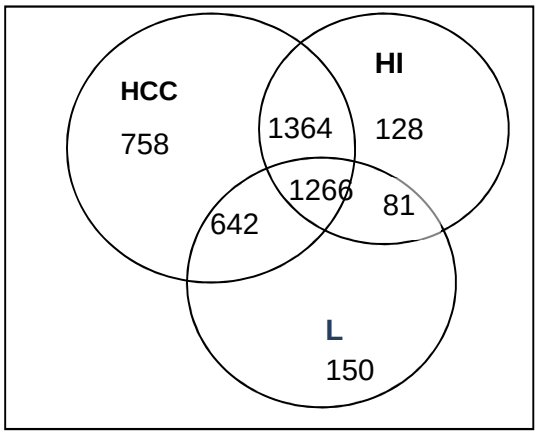
* interaction.

Table 3: Estimates of missing cases for 2008 and 2009 according to sex and age estimated from notified cases and corresponding completeness of the Registry of Lille and its area at the time of national data centralization

Model	Properties		Estimates				
	df	p-value[a]	Missing cases N (95% CI)	Completeness *Rate (%) (95% CI)*	Missing cases N (95% CI)	Completeness *Rate (%) (95% CI)*	Completeness *Rate (%) (95% CI)*
Men			<65 years old		≥65 years old		Overall
HI L HCC Age HI*HCC L*HCC Age*L Age*HI Age*HCC	4	0.61	63.8 (44.3-91.9)	*97.9 (96.8-99.0)*	191.6 (137.9-266.1)	*93.8 (91.4-96.2)*	*95.7 (94.3-97.2)*
Women			<52 years old		≥ 52 years old		Overall
HI L HCC Age HI*HCC L*HCC Age*L Age*HI	5	0.16	33.4 (21.1-52.7)	*96.6 (94.8-98.4)*	201.9 (130.6-312.1)	*94.2 (91.3-97.1)*	*94.8 (92.6-97.0)*

HI: health insurance services; L: pathology labs; HCC: health care centres; df: number of degrees of freedom; CI: Confidence interval.
[a] p-value of the goodness-of-fit test.
* interaction (all are significant).

Figure1: Notifications of cancers to the Registry of Lille and its area by the three sources: health insurance services (HI), pathology labs (L), health care centres (HCC) - Top: men, bottom: women



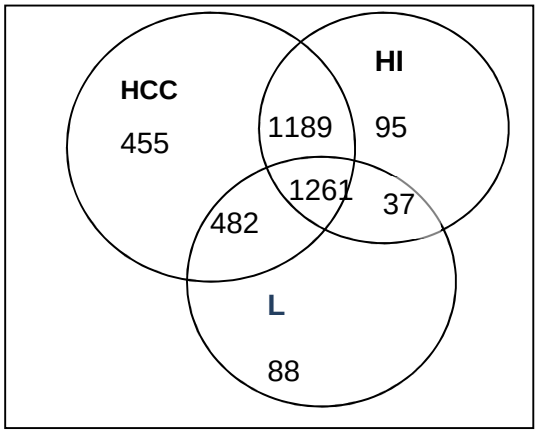


Figure1: Notifications of cancers to the Register of Lille and its area by the three sources: health insurance services (HI), pathology labs (L), health care centres (HCC) - Top: men, bottom : women