

Research on occupational diseases in the absence of occupational data: a mixed-method study among cancer registries of Western Switzerland

Ekaterina Plys^a, Nicolas Bovio^a, Patrick Arveux^b, Yvan Bergeron^c, Jean-Luc Bulliard^{abd}, Nadia Elia^e, Evelyne Fournier^f, Isabelle Konzelmann^g, Manuela Maspoli^h, Elisabetta Rapiti Aylward^f, Irina Guseva Canu^a

^a Centre for Primary Care and Public Health (Unisanté), University of Lausanne, Switzerland

^b Vaud Cancer Registry, Unisanté, Lausanne, Switzerland

^c Fribourg Cancer Registry, Switzerland

^d Neuchâtel and Jura Cancer Registry, Neuchâtel, Switzerland

^e Division of Anaesthesiology, University Hospital of Geneva, Switzerland

^f Geneva Cancer Registry, University of Geneva, Switzerland

^g Valais Cancer Registry, Sion, Switzerland

Summary

The contribution of occupation-related diseases to the global burden of disease is greatly underestimated, mainly due to a shortage of occupational exposure data. This problem is particularly salient in Switzerland, where no estimates of occupation-related disease burden exist, even for the well-recognised occupational cancers, such as malignant pleural mesothelioma and lung cancer.

To overcome this situation, we launched a research project “Examining Cancers and Labour Indicators to assess the Burden” (ExCaLIBur). Within this project, we aimed to assess the need for and quality (i.e., completeness, accuracy and precision) of occupation registration in all cancer registries of Western Switzerland. We also aimed to find a relevant and feasible strategy to collect this information in the future.

We applied a mixed research method. We observed that, independently of the level of precision (5-3-2-1-digit aggregation level), the accuracy was lesser in the registries that were able to actively search and verify occupational information. Overall, the distinction of occupations based on the 3-digit code presents an acceptable compromise in terms of precision. Having such occupations registered in all, or most, Swiss cancer registries routinely would obviously be valuable for epidemiological surveillance of occupational cancers in Switzerland. However, it seems less obvious how these data could fulfill the research objectives, since a better precision than 3-digit occupational coding is challenging to achieve.

Currently, the collection of occupational data by the Swiss cancer registries remains feasible in the frame of specific research projects on occupational cancers. However, available data sources, as well as lack of financial and human resources, will continue to affect quality of the collected occupation data. Therefore, the usage of the standard-

ised questionnaire retracing the individual occupational history to enable further assessment of individual exposure to potential occupational hazards is recommended. However, this approach will disable the Swiss registries to insuring their epidemiological surveillance mission with respect to occupational cancers in Switzerland, for which national statistics remain limited.

Introduction

Occupation is an essential component of adult life and a major determinant of health and healthy ageing [1]. Moreover, some occupational exposures also affect the offspring health [2, 3]. The contribution of occupation-related diseases to the global burden of disease is greatly underestimated, mainly due to a shortage of occupational exposure data [4]. This problem is particularly salient in Switzerland, where no estimates of occupation-related disease burden exist, even for the well-recognised occupational cancers, such as malignant pleural mesothelioma and lung cancer [5, 6]. The reasons for this include: a low interest for occupational medicine [7, 8], little research in the field of occupational health [9], and lack of accessible information on occupation and occupational exposure history in the medico-administrative databases [10]. Apart from statistics produced by the Federal Statistical Office (FSO) from the Swiss Health Surveys and the Swiss Labour Force Survey, Switzerland does not have a systematic data collection strategy to address the issues of occupational exposures and diseases. This hampers aetiological research development, for which occupational history (also called *curriculum laboris*) of each individual should be assessed prospectively or retrospectively, by listing all occupations, their duration and calendar period over their entire career.

Swiss cancer registries have initiated some research on occupation-related cancers and collected information on patients' occupations. However, the latter mainly served as

Correspondence:

Dr Ekaterina Plys
Centre for Primary Care
and Public Health (Unisanté)
Route de Berne 113
CH-1010 Lausanne
Ekaterina.Plys[at]unisanté.ch

an adjustment variable or for investigating socioeconomic differences in cancer incidence, in stage at cancer diagnosis and in cancer survival [11, 12] from public health and healthcare perspectives. The study by Bouchardy et al. [13], published 20 years ago, constitutes the only reference on occupational cancer in Switzerland. It is worth noting that this study provides no cancer incidence rates by occupation because general population data were not available. A case-control design was thus applied, using the patients with the cancer of interest as cases, and patients with all other cancers as controls. Moreover, this study was limited to male cancer cases which occurred during the period 1980–1993. This study suggested increased cancer risk among farmers, butchers, leather and fur workers, founders, electricians, machinists and professional drivers, carpenters, cabinet-makers, hotel and restaurant employees, hairdressers, and chemists. The authors recognised the questionable quality of their occupation information; the occupations were unknown for 23% of cases among men aged 25 to 64 years and for an even greater proportion among men aged over 65 years [13]. However, no further research on occupational cancer has been conducted since except one on skin cancer in relation with solar ultraviolet light exposure [14, 15]. Moreover, female workers, ageing workers and young workers remained neglected with respect to their risk of occupational and occupation-related cancer, although these groups are considered "particularly sensitive risk groups", because of differences in metabolism and pre-existing health problems — including those caused by work [16, 17].

To overcome this situation, we launched a research project entitled "Examining Cancers and Labour Indicators to assess the Burden" (ExCaLIBur) in order to: (1) produce sex- and age-specific cancer risk and survival estimates by occupation and by economic activity in Switzerland; (2) assess the need for and quality of occupation registration in Swiss registries. This second part of the project aims at evaluating the completeness, accuracy and precision of occupations documented in Swiss cancer registries and the relevance of collecting this information for future occupational epidemiology research. The need of such an evaluation was urged by the recent Cancer Registration Act on the compulsory registration of oncological diseases in Switzerland, according to which occupation is not included in the national cancer data structure [18, 19]. Before this Act came in force (1 January 2020), the perimeter of data registration in cancer registries was more flexible, allowing collection of occupational variables and other variables deemed relevant for epidemiological research.

Because cancer registries have access to different sources of information and rely on different procedures to record occupation, we hypothesised that the sources and procedures for collecting data on occupation would lead to the more accurate coding of occupation as compared with the occupation coded by the FSO in the Swiss National Cohort (SNC) [20].

To investigate this hypothesis and understand the sources/reasons of practice and potential heterogeneity between registries, as well as to consider ways to improve the access to and the quality of occupation variable, we conducted a mixed method research [21] in collaboration with the cancer registries of Western Switzerland. This report pre-

sents the results of combined quantitative and qualitative analyses, including focus group and mapping approaches [22].

Methods

We used mixed method because it allowed us to combine quantitative and qualitative approaches. With quantitative analysis, we estimated the quality of the data collected by the registries (i.e., percentage of missing values) and evaluated agreement of occupational information documented in the registries and in the FSO. These results were discussed using a qualitative approach and sensible explanations for these results were proposed. Moreover, the qualitative approach enabled us to analyse the current strategies of occupational data collection and coding, and to discuss feasible strategies that could be used in the future. The combination of the methods also enabled us to take into consideration the opinions and practices of all registries regardless of the use of their data in the main statistical analysis. The first step in our study was collection of the data from the registries and the FSO. Next, the data were linked and quantitative analyses were performed. Finally, we organised a focus group with representatives of the five cancer registries of Western Switzerland to discuss the results, current occupational data collection and coding practices used and options for further data collection.

Data sources

For the quantitative analysis, we used two sources of data: the FSO, which provided access to the SNC, and the cancer registries of Geneva, Neuchâtel-Jura, Valais, and Vaud. The registry of Fribourg did not register occupation of their patients and was excluded. The data collected by FSO were used for comparison, because these data come from reliable sources and the FSO can guarantee the veracity of its data [23].

The SNC is a national longitudinal research platform for the entire resident population of Switzerland. It includes the records of the 1990 and 2000 Swiss censuses, linked to mortality, life birth and emigration records up to 2015, using a combination of deterministic and probabilistic methods [20]. Because federal censuses were mandatory, the SNC covered 98.6% of Swiss population [24]. The occupations recorded in 2000 were coded by the FSO using the Swiss Standard Classification of Occupations 2000 (NSP 2000), the classification most used in Switzerland nowadays. Therefore, the SNC is the only available data source that is relevant to serve as an external reference comparator for occupation for this study purposes.

Occupational data from all cancer cases from the four Cancer Registries of Western Switzerland over the period 1990–2014 were extracted by each registry and sent to Unisanté, where they were harmonised and centralised into a database. The linkage between these data and the SNC data was then established by the Institute of Social and Preventive Medicine based on a probabilistic linkage procedure.

Besides, each registry provided the description of their practices of occupational data collection and coding, the data sources used, the type of collection (i.e., active versus passive; systematic versus occasional), their quality con-

trol procedures, the resources available as well as the challenges related to these practices.

Study sample

The study sample included records of all cancers diagnosed between 1996 and 2004 in four cancer registries that collected occupational data, matched with the SNC. This selection was made in order to enable the comparison of the occupations for this calendar period with occupations declared in the 2000 federal census, under the hypothesis that occupation remained constant during this period [25].

Quantitative component

Firstly, we conducted descriptive analyses of the registries' data, aiming to estimate the proportion of missing values (empty fields) and potentially inaccurate information such as "non-defined occupation" or "no information". We were also interested in the proportion of records coded as "retired" and "housewife". Although these statuses do not enable us to know individuals' previous occupation, if any, they inform us that these individuals are not currently exposed to occupational carcinogens.

Secondly, we compared the occupational information coding agreement and precision between occupation coded in the registries and in the SNC. Only occupations coded using NSP 2000 were used for this comparison. Thus, our analysis included data from the registries of Neuchâtel-Jura, Vaud, and Valais. Since the Geneva registry used the Classification of the Swiss Federal census 1979, their data were not included in the analysis. The comparison between the registry and SNC codes of occupation was performed accounting for the precision of occupation coding in each registry. For this, we stratified the analysis by the NSP 2000 aggregation level, using 5, 3, 2 and 1-digit levels of aggregation, with the 5-digit level being the most detailed coding (i.e., distinguishing 383 occupations) and the 1-digit level the less detailed one (with 9 occupation groups only). When the occupation was coded at a more aggregated level, we completed the code by replacing the missing digit(s) with 0.

If the occupation code was identical between the registry and the SNC, we categorised it under "Same occupation". If the code was different between the two data sources, it was classified as "Mismatch". When the information was missing from both the registry and the SNC, we assigned the case to "No information". Finally, we also compared the agreement of coding of retired patients within each registry, using "Retired" code in both sources. All statistical analyses were stratified by registry and performed using STATA statistical software, version 16 (StataCorp LP; TX, USA).

Qualitative component

The qualitative component of the study consisted of a focus group combined with a mapping approach. The focus group aimed to discuss the results of the quantitative analysis and better possibilities of occupational data collection. The focus group took place online and included representatives of the five cancer registries. It lasted three hours. First, the registries presented their strategies of occupational data collection and coding. Next, the results of

the quantitative analyses were presented. During the discussion that followed, the cancer registry representatives provided possible explanations for the observed results, based on their registry-specific experience. Then, the participants were invited to suggest potential solutions allowing effective and efficient occupational data collection in the context of the new Swiss Cancer Registration Act. One moderator (IGC) and one observer (NB) from the project team managed the focus group, while another observer (EP) was taking notes. The moderator iteratively summarised the suggested solutions, which were (re)debated in several rounds of discussion until a consensus on their respective relevance could be reached among the registry representatives.

Finally, the feasibility of the most relevant strategies was assessed by conducting a mapping study [22]. The latter consisted of an iterative review of potential sources of occupational data aimed to reveal qualitative descriptors and linkages between occupation-related data and information flow on the federal level in Switzerland.

This study was approved by the Vaud ethics committee for research (N° 2018-02077), and received funding from the Swiss Cancer Research.

Results

Registry-specific practices for occupational data collection and coding

Table 1 summarises the registries' practices in collection and coding occupation over time.

The cancer registries of Western Switzerland have long recognised the importance of occupational data collection not only for occupational epidemiology but also for the development of effective prevention strategies. Four registries out of five collected these data and considered it in accordance with their mission of cancer monitoring in Switzerland. However, some registries (Fribourg and Valais) believe that occupational data serve mostly research rather than surveillance purposes and are less invested. Fribourg registry has not collected occupational data at all. Most registries encounter difficulties in collection of the information about patients' occupation due to a lack of financial and human resources. The registries of Geneva, Valais and Vaud mentioned that they could not afford to hire trained people to collect information on occupation, and the registry of Neuchâtel-Jura is the only registry that had an employee partly in charge of occupational data collection and coding for 12 years.

Identification of the patients' socioeconomic status was one of the main objectives of occupational data collection. Since the 2020 legislation on the compulsory registration of oncological diseases, only the registry of Neuchâtel-Jura has obtained cantonal permission to continue occupational data collection. The other registries had to stop it. All the registries used medical records and mortality data provided by the FSO as a data source about occupation. Nevertheless, they also referred to other sources such as oncological reports, cantonal or municipal population office, or cantonal population registry book.

The registries of Neuchâtel-Jura and, until 2015, Geneva actively searched and verified occupations. Moreover, the

registry of Neuchâtel-Jura assessed the economic activity branch or industry of their patients' employer, based on the company name and accounted for it in their coding of occupations. The registries of Geneva and Neuchâtel-Jura also searched and documented the last occupation of the retired patients. Further, the registry of Neuchâtel-Jura compared their documented occupation with the occupation provided by the FSO on the death certificate and applied changes if necessary. The occupational data in both registries were collected and coded by a dedicated employee who performed this activity for many years. As primary sources for the occupation they both used data of the population registry and those provided in medical records. Conversely, the registries of Vaud and Valais collected only the occupational information available without any active data search or verification. Regarding data quality, the collected occupational data were occasionally checked by a physician at the registry of Neuchâtel-Jura, while the registry of Vaud performed checks only for data used in research studies.

Currently, all the registries except Geneva code occupation using the Swiss classification of occupations adopted in 2000. In the Geneva registry, occupations were coded based on the classification of occupations of the Swiss Federal Census of 1979. Nevertheless, several other classifications of occupations were used by the registries in the past (table 1).

Registries acknowledged that their occupational data might be incomplete and inaccurate as their sources of information are not always reliable. For instance, several registries reported lack of data on women's occupations. The registries also expressed their concern about being able to collect only the last occupation of their patients, which is insufficient for research in occupational epidemiology.

Description of registries' occupational data

In total, the four registries registered 97,571 cancer cases between 1996 and 2004. All data were used for descriptive statistics.

Table 2 shows that the proportion of missing values in registries' data is small, ranging between 0% and 5.2%. We observed the same result for housewives, for whom the proportion of missing values ranged between 0% and 1.6%. Records coded as "no information" were particularly rare in registries of Geneva and Neuchâtel-Jura (0% and 3.7%, respectively). However, they accounted for 59.2% of the data provided by Valais. Data presented by Neuchâtel-Jura and Vaud contained few records with undefined occupation (3% and 2.5%, respectively), whereas they accounted for 12.6% and 14.5% of records in Valais and Geneva, respectively.

Table 1:
Coding procedures of the occupational information among the five cancer registries of western Switzerland.

Registry	Source of information (period)	Classification type and objective of data collection	Quality of the collected data
Valais	Cantonal office of population (1989-2011) ^P / Medical records (2012-2019) ^P / Mortality data from the FSO (2014-2019) ^P	Classification: Classification of the ASRT; NSP 2000 Objective: No particular objective	Data occasionally collected between 1989 and 2019; Data collection stopped in 2020 Quality check: None
Neuchâtel-Jura	Civil registry (Social assistance service) (1974-1995) ^a / Medical records (from 1996) ^a / Oncological reports (from 2005) ^a / Mortality data from the FSO (from 2016) ^a	Classification: 1974-1991: only 3 types of occupations (worker, employee/middle-class, executive); 1992-1997: Classification of the Swiss Federal census 1980; 1998-2016: NSP 1990; From 2016: NSP 2000; In 2016 all the documented cases was re-coded in accordance with the NSP 2000 Objectives: Evaluate socio-economic status of the patients; Use occupational data in epidemiological studies	Data systematically collected from 1974; Data collection still possible after 2020 Specificity of data coding: If the patient is retired but his/her occupation is known, the occupation is coded; Comparison of the collected data with those provided by the FSO; modifications are made if necessary (from 2016) Quality check: Occasional check done by a physician
Vaud	Municipalities' population registry (till 2014) ^P / Medical records (till 2017) ^P / Mortality data from the FSO ^P	Classification: NSP 2000 Objective: Evaluate risk of cancer incidence depending on socio-economic and occupational status of the patients	Data collection between 1974 and 2017; Data systematically collected till 2010; Data collection stopped in 2020 Quality check: Occasional (data used for research studies)
Geneva	Cantonal office of population (from 1970) ^P / Mortality data from the FSO (if occupation field is empty at the time of death) / Population registry book of the Geneva canton (1970-1999) ^a / Medical records (from 1970) ^P	Classification: Classification of the Swiss Federal census 1979 Objective: Evaluate socio-economic status of the patients	Data systematically collected between 1980 and 2015; Data occasionally collected between 2015 and 2020; Data collection stopped in 2020 Specificity of data coding: If the occupation is unknown, the case is coded as « No occupation » or « Occupation is unknown »; If the patient is retired at the time of diagnosis, the last occupation is coded Quality check: None

In bold: main source of the occupational information; ^a active data collection (some specific steps are undertaken to collect data, for instance, search for the occupation in a population registry book, search of the information on exposures to environmental carcinogens, check of the available information); ^P passive data collection (coding of the occupational information as it was transmitted).

ASRT: Swiss association of cancer registries; FSO: Federal Statistical Office; ISCO 88 (COM): International standard classification of occupations; NSP 1990: Swiss standard classification of occupations 1990; NSP 2000: Swiss standard classification of occupations 2000.

Comparison of occupation codes between registries and the SNC

Table 3 and Supplementary figure S1 (appendix) show that differences in agreement varies depending on the levels of precision in the occupation coding. We observed that the finer the level of aggregation, the lower the percentage of concordance. For example, in the registry of Neuchâtel-Jura 11% of the records had the same occupational code at 1-digit, while this number decreased to 7% when occupation was coded at 5-digits. Based on these findings, it seems that using the 3-digit code can be an acceptable compromise. Therefore, all subsequent results are presented at the 3-digit level.

We identified that occupational information differed by at least 50% between the two data sources. Data from the registry of Neuchâtel-Jura had the lowest level of mismatch (46%) followed by the registry of Vaud (55%). The registry of Valais had 81% mismatch, which was largely due to the fact it coded retired patients as “no information”.

We aimed at comparing the agreement of occupation coding by the registers with that by the FSO. We considered coding concordant when the occupational information was missing or when the patient was coded “retired” in both sources. Thus, most often occupations matched in the data of the Vaud registry (11%), followed by Neuchâtel-Jura (9%), and Valais (5%). However, when we took into account the percentage of mismatch or data on retired patients, we see that the most accurate information was collected by the registry of Neuchâtel-Jura (table 3). Neuchâtel-Jura had 46% of mismatched data, whereas Vaud and Valais had 55% and 81% of mismatched data, respectively. Regarding accuracy of coding for retired status, Neuchâtel-Jura had 44% of match with SNC data, whereas Vaud and Valais had 29% and 0%.

Strategies for occupational data collection defined within the focus group

During the focus group meeting, the participants commented on the results of quantitative analysis. The observed dif-

Table 2: Completeness of occupational information in cancer registries of Western Switzerland between 1996 and 2004.

Registry	Number of recorded cases	Non-defined occupation ^a		Retired		Housewife		No information ^b		Missing values ^c	
		n	%	n	%	n	%	n	%	n	%
Neuchâtel-Jura	9662	289	3	5711	59.1	7	0.1	360	3.7	1	0
Vaud	45,264	110	2.5	18,765	41.5	4	0	8696	19.2	22	0.1
Valais	14,832	1872	12.6	0	0	239	1.6	8774	59.2	722	4.9
Geneva	27,813	4035	14.5	–	–	–	–	11	0	1452	5.2
Total	97,571	6306	6.5	24,476	25.1	250	0.3	17,841	18.3	2197	2.3

^a Unknown or unclassified occupations; ^b no information was collected about the occupation; ^c empty field.

Table 3: Comparison of coded occupational information in the cancer registries of Western Switzerland between 1996 and 2004 with that of the Swiss National Cohort in the 2000 federal census.

Registry *	Neuchâtel-Jura		Vaud		Valais	
	n	%	n	%	n	%
<i>NSP 2000 1 digit</i>						
Mismatch	3391	44	19,356	52	9156	79
Same occupation	870	11	4937	13	745	6
Retired	3389	44	10,658	29	0	0
No information	114	1	2003	5	1693	15
<i>NSP 2000 2 digits</i>						
Mismatch	3516	45	19,996	54	9250	80
Same occupation	745	10	4297	12	651	6
Retired	3389	44	10,658	29	0	0
No information	114	1	2003	5	1693	15
<i>NSP 2000 3 digits</i>						
Mismatch	3580	46	20,354	55	9338	81
Same occupation	681	9	3939	11	563	5
Retired	3389	44	10,658	29	0	0
No information	114	1	2003	5	1693	15
<i>NSP 2000 5 digits</i>						
Mismatch	3732	48	21,236	57	9517	82
Same occupation	529	7	3057	8	384	3
Retired	3389	44	10,658	29	0	0
No information	114	1	2003	5	1693	15
Total	7764	100	36,954	100	11,594	100

NSP 2000: Swiss standard classification of occupations 2000.

* Only occupations coded using NSP 2000 were used in this analysis. Since the registry of Geneva coded their occupations according to the classification of the Swiss federal census 1979, their data were excluded.

ferences in data quality between the registries seems determined by the data collection practices. Thus, the registry of Neuchâtel-Jura systematically documented if their patients were retired and actively searched for information about patients' occupation, whereas those of Valais and Vaud did not. This situation also reflects differences in availability of human and financial resources, and stresses the importance of providing the registries with reliable occupational information that can be easily coded. The high percentage of occupations matched with the FSO coded occupations observed in the Vaud registry, explained by the fact that the latter used data provided by the FSO as main source of occupation, is a reliable illustration of this.

Regarding the strategies aiming to improve the collection of occupational information on patients, cancer registries representatives agreed unanimously on the usefulness of data directly transmitted by a competent institution, such as the Cantonal compensation office. This office is in charge of retirement pension payment, which is calculated based on individual's occupational experience. This option was judged optimal as the registries have the identification number of the old-age and survivors' insurance (OASI) of their patients, which facilitates the linkage of data from different sources. The registries also suggested some steps aiming to provide all of them with permission to collect patients' occupation through this source and proposed submitting an appropriate request to competent authorities. However, this strategy would be successful only if additional resources are provided for insuring a sufficient quality standard of occupational information registered. In fact, the information on occupation gathered by the Cantonal population offices is self-declared and might be inaccurate or imprecise. The registries also mentioned that if none of the proposed strategies is accepted, it would still be possible to continue collecting occupational data using occupational history questionnaires within the framework of specific research studies. The latter can be done in two different ways: either the registries collect the occupation of all newly registered patients diagnosed with the studied cancer during a limited period (e.g., two years) or they participate in the occupational history data collection by sending a standardised questionnaire to patients or their families. The latter solution seems more appropriate, as using different methods of occupation and exposure assessment in cancer cases and controls would introduce a differential exposure misclassification bias.

Discussion

The first aim of this study was to evaluate the agreement of occupational information documented in Swiss cancer registries compared with the FSO data. We hypothesised that the sources and procedures for collecting data on occupation would affect the agreement or coding of occupation as compared with the occupation in the SNC. Our results allow the conclusion that discrepancies in occupational coding are related to the different procedures of data collection used by the registries. Thus, unsurprisingly, we observed that, independently of the level of precision (5-3-2-1-digit aggregation level), the percentage of mismatches was less in the registries that were able to actively search and verify occupational information.

Our results suggest that using the 3-digit code can be a compromise between accuracy and precision. In fact, it allows clear identification of the occupation of the patient and avoids excluding those patients whose occupation was coded slightly differently as compared with the occupation code at 5-digit precision assigned by the FSO in the SNC. For example, in the NSP 2000, at the 1-digit level, the code 8 corresponds to a large group of occupations related to health, teaching, culture and science. At the 2-digit level, it allows smaller groups such as "occupations related to media" (NSP 2000 code 81) or "health professionals" (NSP 2000 code 86) to be distinguished. At the 3-digit level, the occupation is even more specific, distinguishing for instance "occupations related to human medicine and pharmaceuticals" (NSP 2000 code 861) among the health professionals. Finally, the 5-digit level distinguishes specific occupations such as physician (861.01) or pharmacist (861.03) among the latter.

Our second aim was to evaluate the relevance of the further collection of occupational information by the registries. We identified many current important barriers to further occupational data collection. After the adoption of the Cancer Registration Act, Swiss cancer registries must ask for cantonal legislation to collect occupational data of their patients. Currently, only the registry of Neuchâtel-Jura has obtained cantonal legal authorisation to continue collecting these data. However, even if the appropriate authorisations are obtained, available data sources, as well as lack of financial and human resources, will continue affecting the quality of the collected occupational information.

Having occupation registered in all, or most, Swiss cancer registries routinely would obviously be valuable for epidemiological surveillance of occupational cancers in Switzerland. Indeed, the registries of Geneva and Neuchâtel-Jura searched and coded for occupations of the retired patients, whereas the FSO provided only the individual's status "retired/annuitant". However, it seems less obvious how these data could fulfil the research objectives, since a better precision than 3-digit occupational coding is challenging. Although a 3-digit occupation code is relevant for surveillance purposes, it would not suffice when a linkage with an exposure database (e.g., job exposure matrix for night shift work [26, 27], lung carcinogens [28], or radioactive materials [29]) based on a 4 or 5-digit occupation classification is necessary. Given these findings, the relevance of further collection of occupations for research purposes by cancer registries appears questionable. Indeed, for research purposes, occupational history needs to be reconstructed very precisely. Thus, use of centrally collected data may be of great importance.

Two options emerged from our focus group investigation: [1] use of OASI data and [2] request to the competent authorities in order to obtain permission for all the cancer registries to collect occupational data. However, after our discussion none of these options seems viable. The data collected by OASI would be a potentially relevant source as it provides a unique personal identifier. In fact, such identifiers are widely used in Scandinavian countries [30–32] and, recently, in France [33] including for research purposes. In Norway, unique national identity numbers were used to link data from medical, educational and other registries in order to predict unemployment of young

people [32]. Unfortunately, OASI does not collect information on either the occupation of the insured person or about his/her occupational activity branch or working conditions, but only the enterprise number and the worker's salary for calculation of the contribution amount. It is thus impossible to retrace the patients' occupational history using OASI data. Nevertheless, the FSO is able to link the enterprise number to the Swiss enterprises registry and assess its economic activity branch. Using the latter along with the salary information from the OASI it could be possible to deduce the socioeconomic status of the worker in each enterprise and, perhaps, the occupation, although imprecisely. Because the OASI data cannot be used, the second strategy was also abandoned.

Another option consists in linking data on occupation collected by FSO with data collected by the registries. However, this option was not obvious for the registries and was not discussed with the focus group. More consultations with representatives of the registries and probably with the FSO are needed to determine precisely the feasibility of these options and their interest for the registries and occupational cancer research.

Thus, the routine collection of occupational data given the existing sources, the current legislative context, resources available and our findings, seems challenging and inefficient. Consequently, such a data collection should be limited to specific research studies and conducted according to a research plan approved by relevant ethics committees. In such a research context, the strategy of collection and coding occupation for a specific cancer location during a limited calendar period by the cancer registries also seems suboptimal. In fact, it would only inform on the occupation in incident cancer cases but not in the controls (e.g., patients with other cancer locations or healthy people) or external comparison group (e.g., general (sub-)population). Retrieving their occupation would require additional, potentially different investigation; consequently, such a strategy is prone to a high risk of introducing different exposure classification bias in the research results. Conversely, the retrospective reconstruction of a *curriculum laboris* using a standardised questionnaire provided to both cancer patients and controls remains a relevant and feasible solution for continuing research on occupational cancer in Switzerland. Indeed, it is the only possible strategy in the current context. This approach was shown to be reliable in many epidemiological studies and allowed assessment of the exposure-response relationship for many occupational carcinogens in national and international studies [34–36]. The coding of occupations during the entire career is a burdensome and time-consuming task, and constitutes the main limitation of this approach. However, the automation of the coding of occupations thanks to a new software “PROCEDURE” could alleviate this limitation [38]. The second drawback is that such a questionnaire significantly elongates the study questionnaire and the time required for its completion, which could bother the participants and lead to dropouts. To remedy this issue, a simplified questionnaire on occupational history could be used, such as one currently tested within the Swiss Health Study Pilot Project [39], as well as more interactive on-line questionnaires that could be filled in using a smartphone or a tablet. Indeed, the collaboration of Swiss cancer registries in the

administration of such a questionnaire to their patients is an additional key-point in this strategy.

Study strengths and limitations

This study was the first to examine the quality of occupational data using a systemic research protocol applied to five population-based Swiss cancer registries. The study findings confirmed the research hypothesis that the quality of data estimated as completeness and precision of registries' occupational data depends on the sources used for data collection and the methodological rigour of registration and coding procedures, the latter being determined by the resources available. The original study findings were timely, given the new regulation on cancer registration in Switzerland, which came into operation in January 2020.

The use of a probabilistic linkage procedure implies that for some individuals cancer registries and SNC records could not be linked successfully. However, according to Schmidlin et al., only 5.4% of records could not be linked using stringent probabilistic linkage and the unlinked death records have little effect on relative mortality and reflect well the official rates [40].

The study was limited to all cancer registries of Western Switzerland, raising concern about generalisability of the findings to all the other Swiss registries. Nevertheless, we covered most administrative sources available and these sources are largely comparable across cantons. Therefore, the diversity observed in data collection procedures in our study is likely representative of the procedures used in a broader sample of Swiss cancer registries. The unavailability of occupations coded using NSP 2000 in the Geneva registry and lack of occupational data in the Fribourg registry precluded analysis of the accuracy and precision of these registries. However, it is unlikely that inclusion of them in the analysis would change the study results and conclusions.

The strategies discussed within the focus group and investigated by mapping are federal-level strategies. Consequently, they are relevant for all Swiss registries of diseases. The conclusions regarding the most relevant approach to collect patients' occupation and promote the research on occupational diseases and occupational cancer risks are therefore generalisable to all health outcomes in the country, given its particularities in data availability, flow, interoperability and linkages.

Conclusion

This study is the first investigating the quality of occupational data collected by Swiss cancer registries. The usage of a mixed method protocol added to its originality. The study confirmed the research hypothesis that the quality of occupational data estimated in terms of data completeness, accuracy and precision, depends on the sources used by the registry for data collection and the methodological rigour of registration and coding procedures, the latter being strongly determined by the local legislation and resources available. Given that currently cancer registries cannot collect occupational data without appropriate cantonal legislation, the simplest solution is to retrace occupational history in the frame of specific studies. However,

er, available data sources, as well as lack of financial and human resources, will continue affecting quality of the occupational information collected by the registries. Therefore, the usage of a standardised questionnaire retracing individual occupational history to enable further assessment of individual exposure to potential occupational carcinogens and hazards will be more relevant than using routinely available data. However, this approach will leave the Swiss registries unable to fulfil their epidemiological surveillance mission with respect to occupational cancers, for which national statistics remain limited. The prevention and consequently the surveillance of occupational diseases, including occupational cancers, is one of the missions of the public health authorities, in the same sense that occupational health is a part of public health [9].

Financial disclosure

This work was supported by Grant No KFS-4699-02-2019 from the Swiss Cancer Research Foundation.

Potential competing interests

All authors have completed and submitted the International Committee of Medical Journal Editors form for disclosure of potential conflicts of interest. No potential conflict of interest was disclosed.

References

- Global GB; GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018 Nov;392(10159):1923-94. [http://dx.doi.org/10.1016/S0140-6736\(18\)32225-6](http://dx.doi.org/10.1016/S0140-6736(18)32225-6). *PubMed*. 1474-547X
- Adachi S, Sawaki J, Tokuda N, Tanaka H, Sawai H, Takeshima Y, et al.; Japan Environment and Children's Study Group. Paternal occupational exposure to chemicals and secondary sex ratio: results from the Japan Environment and Children's Study. *Lancet Planet Health*. 2019 Dec;3(12):e529-38. [http://dx.doi.org/10.1016/S2542-5196\(19\)30239-6](http://dx.doi.org/10.1016/S2542-5196(19)30239-6). *PubMed*. 2542-5196
- Pape K, Svanes C, Sejbæk CS, Malinowski A, Benediksdottir B, Forsberg B, et al. Parental occupational exposure pre- and post-conception and development of asthma in offspring. *Int J Epidemiol*. 2020. *PubMed*. 0300-5771
- Loomis D. Estimating the global burden of disease from occupational exposures. *Occup Environ Med*. 2020 Mar;77(3):131-2. <http://dx.doi.org/10.1136/oemed-2019-106349>. *PubMed*. 1470-7926
- Bovio N, Richardson DB, Guseva Canu I. Sex-specific risks and trends in lung cancer mortality across occupations and economic activities in Switzerland (1990-2014). *Occup Environ Med*. 2020 Aug;77(8):540-8. <http://dx.doi.org/10.1136/oemed-2019-106356>. *PubMed*. 1470-7926
- Bovio N, Wild P, Guseva Canu I; Swiss National Cohort (SNC). Lung Cancer Mortality in the Swiss Working Population: The Effect of Occupational and Non-Occupational Factors. *J Occup Environ Med*. 2021 Dec;63(12):1029-36. <http://dx.doi.org/10.1097/JOM.0000000000002302>. *PubMed*. 1536-5948
- Danuser B. Comment soigner la santé au travail? REISO. 2014 Jun;2014:1-4.
- Graczyk H, François M, Krief P, Guseva Canu I. The role of the Swiss list of occupational diseases in the protection of workers' health. *Swiss Med Wkly*. 2021 Jul;151:w20538. <http://dx.doi.org/10.4414/smw.2021.20538>. *PubMed*. 1424-3997
- Guseva Canu I, François M, Graczyk H, Vernez D. Healthy worker, healthy citizen: the place of occupational health within public health research in Switzerland. *Int J Public Health*. 2020 Jan;65(1):111-20. <http://dx.doi.org/10.1007/s00038-019-01245-w>. *PubMed*. 1661-8564
- OECD. Health Data Governance: Privacy, Monitoring and Research. Paris; 2015.
- Feller A, Schmidlin K, Bordoni A, Bouchardy C, Bulliard JL, Camey B, et al.; SNC and the NICER workinggroup. Socioeconomic and demographic disparities in breast cancer stage at presentation and survival: A Swiss population-based study. *Int J Cancer*. 2017 Oct;141(8):1529-39. <http://dx.doi.org/10.1002/ijc.30856>. *PubMed*. 1097-0215
- Feller A, Schmidlin K, Bordoni A, Bouchardy C, Bulliard JL, Camey B, et al.; SNC and the NICER Working Group. Socioeconomic and demographic inequalities in stage at diagnosis and survival among colorectal cancer patients: evidence from a Swiss population-based study. *Cancer Med*. 2018 Apr;7(4):1498-510. <http://dx.doi.org/10.1002/cam4.1385>. *PubMed*. 2045-7634
- Bouchardy C, Schuler G, Minder C, Hotz P, Bousquet A, Levi F, et al. Cancer risk by occupation and socioeconomic group among men—a study by the Association of Swiss Cancer Registries. *Scand J Work Environ Health*. 2002;28 Suppl 1:1-88. *PubMed*. 0355-3140
- Backes C, Religi A, Moccozet L, Vuilleumier L, Vernez D, Bulliard JL. Facial exposure to ultraviolet radiation: predicted sun protection effectiveness of various hat styles. *Photodermatol Photoimmunol Photomed*. 2018 Sep;34(5):330-7. <http://dx.doi.org/10.1111/phpp.12388>. *PubMed*. 1600-0781
- Milon A, Bulliard JL, Vuilleumier L, Danuser B, Vernez D. Estimating the contribution of occupational solar ultraviolet exposure to skin cancer. *Br J Dermatol*. 2014 Jan;170(1):157-64. <http://dx.doi.org/10.1111/bjd.12604>. *PubMed*. 1365-2133
- EU-OSHA. Exposure to carcinogens and work-related cancer: A review of assessment methods. Executive summary. Luxembourg: European Agency for Safety and Health at Work (EU-OSHA); 2014.
- EU-OSHA. Priorities for occupational safety and health research in Europe for the years 2013-2020. Summary report. Luxembourg: European Agency for Safety and Health at Work (EU-OSHA); 2014.
- (OFSP) Ofdisp. Législation sur l'enregistrement des cancers Bern, Suisse2021 [updated 28.04.2021. Available from : <https://www.bag.admin.ch/bag/fr/home/gesetze-und-bewilligungen/gesetzgebung/gesetzgebung-mensch-gesundheit/gesetzgebung-krebsregistrierung.html>]
- Bulliard JL, Ducros C, Germann S, Arveux P, Bochud M. Nouvelles exigences légales dans le domaine de l'enregistrement du cancer : opportunités et défis. *Rev Med Suisse*. 2020 Nov;16(713):2099-103. *PubMed*. 1660-9379
- Spoerri A, Zwahlen M, Egger M, Bopp M. The Swiss National Cohort: a unique database for national and international researchers. *Int J Public Health*. 2010 Aug;55(4):239-42. <http://dx.doi.org/10.1007/s00038-010-0160-5>. *PubMed*. 1661-8564
- Johnson RB, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educ Res*. 2004;33(7):14-26. <http://dx.doi.org/10.3102/0013189X033007014>. 0013-189X
- Cooper ID. What is a "mapping study"? *J Med Libr Assoc*. 2016 Jan;104(1):76-8. <http://dx.doi.org/10.3163/1536-5050.104.1.013>. *PubMed*. 1558-9439
- Lorenz H. The Data Innovation Project "NOGAuto". SWISS FEDERAL STATISTICAL OFFICE; 2021.
- SFSO. Methodology report—coverage estimation for the Swiss population census 2000. Swiss Federal Statistical Office Neuchâtel2004 [Available from: <https://www.bfs.admin.ch/bfsstatic/dam/assets/341896/master>]
- Guseva Canu I, Bovio N, Mediouni Z, Bochud M, Wild P; Swiss National Cohort (SNC). Suicide mortality follow-up of the Swiss National Cohort (1990-2014): sex-specific risk estimates by occupational socioeconomic group in working-age population. *Soc Psychiatry Psychiatr Epidemiol*. 2019 Dec;54(12):1483-95. <http://dx.doi.org/10.1007/s00127-019-01728-4>. *PubMed*. 1433-9285
- Sok V, Tvardik N, Cordina E, Pelletan JB, Houot M, Pilorget C, et al. Création d'une matrice emplois-expositions sur le travail à horaires atypiques. *Arch Mal Prof Environ*. 2018;79(3):400. <http://dx.doi.org/10.1016/j.admp.2018.03.427>. 1775-8785
- Fernandez RC, Peters S, Carey RN, Davies MJ, Fritschi L. Assessment of exposure to shiftwork mechanisms in the general population: the development of a new job-exposure matrix. *Occup Environ Med*. 2014 Oct;71(10):723-9. <http://dx.doi.org/10.1136/oemed-2014-102101>. *PubMed*. 1470-7926
- Peters S, Vermeulen R, Portengen L, Olsson A, Kendzia B, Vincent R, et al. SYN-JEM: A Quantitative Job-Exposure Matrix for Five Lung Carcinogens. *Ann Occup Hyg*. 2016 Aug;60(7):795-811. <http://dx.doi.org/10.1093/annhyg/mew034>. *PubMed*. 1475-3162
- Guseva Canu I, Faust S, Kniecjak E, Carles M, Samson E, Laurier D. Estimating historic exposures at the European Gaseous Diffusion Plants. *Int J Hyg Environ Health*. 2013 Jul;216(4):499-507. <http://dx.doi.org/10.1016/j.ijheh.2012.07.002>. *PubMed*. 1618-131X
- Foss L, Gravseth HM, Kristensen P, Clausen B, Mehlum IS, Skyberg K. "Inclusive working life in Norway": a registry-based five-year follow-up study. *J Occup Med Toxicol*. 2013 Jul;8(1):19. <http://dx.doi.org/10.1186/1745-6673-8-19>. *PubMed*. 1745-6673
- Ludvigsson JF, Håberg SE, Knudsen GP, Lafolie P, Zoega H, Sarkkola C, et al. Ethical aspects of registry-based research in the Nordic countries. *Clin Epidemiol*. 2015 Nov;7:491-508. <http://dx.doi.org/10.2147/CLEP.S90589>. *PubMed*. 1179-1349
- Kristensen P, Hanvold TN, Hasting RL, Merkus SL, Hoff R, Mehlum IS. Work participation in young Norwegians: a 19-year follow up in a registry-based life-course cohort. *Scand J Public Health*.

- 2021 Mar;49(2):176–87. <http://dx.doi.org/10.1177/1403494820917507>. PubMed. 1651-1905
33. (Irdes) Idredeédls. HYGIE Project: Information system on daily allowances Paris, France: Irdes; 2021 [Available from: <https://www.irdes.fr/recherche/partenariats/hygie-systeme-d-information-sur-les-indemnitees-journalieres/presentation-de-la-base.html>]
34. Hall AL, Kromhout H, Schüz J, Peters S, Portengen L, Vermeulen R, et al. Laryngeal cancer risks in workers exposed to lung carcinogens: exposure-effect analyses using a Quantitative job exposure matrix. *Epidemiology*. 2020 Jan;31(1):145–54. <http://dx.doi.org/10.1097/EDE.0000000000001120>. PubMed. 1531-5487
35. Guha N, Bouaoun L, Kromhout H, Vermeulen R, Brüning T, Behrens T, et al. Lung cancer risk in painters: results from the SYNERGY pooled case-control study consortium. *Occup Environ Med*. 2021 Apr;78(4):269–78. <http://dx.doi.org/10.1136/oemed-2020-106770>. PubMed. 1470-7926
36. Ge C, Peters S, Olsson A, Portengen L, Schüz J, Almansa J, et al. Respirable Crystalline Silica Exposure, Smoking, and Lung Cancer Subtype Risks. A Pooled Analysis of Case-Control Studies. *Am J Respir Crit Care Med*. 2020 Aug;202(3):412–21. <http://dx.doi.org/10.1164/rccm.201910-1926OC>. PubMed. 1535-4970
37. Ge C, Peters S, Olsson A, Portengen L, Schüz J, Almansa J, et al. Diesel engine exhaust exposure, smoking, and lung cancer subtype risks. A Pooled Exposure-response analysis of 14 case-control studies. *Am J Respir Crit Care Med*. 2020 Aug;202(3):402–11. <http://dx.doi.org/10.1164/rccm.201911-2101OC>. PubMed. 1535-4970
38. Savic N, Bovio N, Gilbert F, Paz J, Guseva Canu I. Procode: A Machine-learning tool to support (re-)coding of free-texts of occupations and industries. *Ann Work Expo Health*. 2021 Jun;wxab037. <http://dx.doi.org/10.1093/annweh/wxab037>. PubMed. 2398-7316
39. Unisanté. Swiss Health Study (SHeS, pilot phase) Lausanne, Switzerland 2021 [Available from: <https://www.unisanté.ch/fr/formation-recherche/recherche/groupe-recherche/swiss-health-study-shes-pilot-phase>]
40. Schmidlin K, Clough-Gorr KM, Spoerri A, Egger M, Zwahlen M; Swiss National Cohort. Impact of unlinked deaths and coding changes on mortality trends in the Swiss National Cohort. *BMC Med Inform Decis Mak*. 2013 Jan;13(13):1. <http://dx.doi.org/10.1186/1472-6947-13-1>. PubMed. 1472-6947

Appendix

Supplementary figure

Figure S1: Proportions of the occupations coded using the same International standard classification of occupations (ISCO) code as in the Swiss National Cohort between 1996 and 2004.

