RESEARCH ARTICLE

Histology-specific standardized incidence ratio improves the estimation of second primary lung cancer risk

Marian Eberl[1*] , Luana F. Tanaka[1] , Klaus Kraywinkel[2] and Stefanie J. Klug[1]

Abstract

Background Lung cancer (LC) survivors are at increased risk for developing a second primary cancer (SPC) compared to the general population. While this risk is particularly high for smoking-related SPCs, the published standardized incidence ratio (SIR) for lung cancer after lung cancer is unexpectedly low in countries that follow international multiple primary (IARC/IACR MP) rules when compared to the USA, where distinct rules are employed. IARC/IACR rules rely on histology-dependent documentation of SPC with the same location as the first cancer and only classify an SPC when tumors present different histology. Thus, SIR might be underestimated in cancer registries using these rules. This study aims to assess whether using histology-specific reference rates for calculating SIR improves risk estimates for second primary lung cancer (SPLC) in LC survivors.

Methods We (i) use the distribution of histologic subtypes of LC in population-based cancer registry data of 11 regional cancer registries from Germany to present evidence that the conventional SIR metric underestimates the actual risk for SPLC in LC survivors in registries that use IARC/IACR MP rules, (ii) present updated risk estimates for SPLC in Germany using a novel method to calculate histological subtype-specific SIRs, and (iii) validate this new method using US SEER (Surveillance, Epidemiology, and End Results Program) data, where different MP rules are applied.

Results The adjusted relative risk for lung cancer survivors in Germany to develop an SPLC was 2.98 (95% CI 2.53–3.49) for females and 1.15 (95% CI 1.03–1.27) for males using the novel histology-specific SIR. When using IARC/IACR MP rules, the conventional SIR underestimates the actual risk for SPLC in LC survivors by approximately 30% for both sexes.

Conclusions Our proposed histology-specific method makes the SIR metric more robust against MP rules and, thus, more suitable for cross-country comparisons.

Keywords Second primary cancer, Lung cancer, Cancer epidemiology, Cancer registry data, Standardized incidence ratio

*Correspondence:
Marian Eberl
marian.eberl@tum.de
[1] TUM School of Medicine and Health, Technical University of Munich, Georg-Brauchle-Ring 56, Munich 80992, Germany
[2] German Centre for Cancer Registry Data, Robert Koch-Institut, Nordufer 20, Berlin 13353, Germany

Background

The International Agency for Research on Cancer (IARC) collects, harmonizes, and publishes data on cancer incidence worldwide. According to recent estimates, 2.2 million lung cancer (LC) cases were newly diagnosed worldwide in 2020, comprising 11.4% of all cancer diagnoses, with an increasing incidence trend among women

Eberl *et al. BMC Medicine* (2024) 22:187

Page 2 of 12

[1, 2]. Lung cancer survivors are at an overall increased risk for developing a second primary cancer (SPC) compared to the general population. Subsequent tumors of the oral cavity, pharynx, digestive, and respiratory organs are of particular importance because a large proportion of survivors are long-time smokers who continue smoking [3–6]. However, for some countries, such as Germany or the UK, published standardized incidence ratios (SIR) for second primary lung cancer (SPLC) after lung cancer are unexpectedly low and even suggest a risk reduction for diseased males compared to the general population (e.g., 0.83 and 2.06 for males and females in Germany [7]; 1.47 in the UK (not stratified by sex) [8]). Data from the USA, in contrast, indicate that lung cancer survivors are three to five times more likely to have another lung tumor than the general population, which has lower smoking rates (SIR = 3.38 for males and 4.85 for females) [9].

An essential difference between the US data with a high risk for SPLC and German data with a much lower risk is that national registries apply different rules defining SPLCs. The German registries, and most other international registries, follow the rules for multiple primary cancers (MP rules) defined by the International Agency for Research on Cancer and the International Association for Cancer Registration (IARC/IACR) [10]. They define that only 1 primary cancer per organ can be diagnosed in a patient's lifetime unless multiple tumors are "histologically different." IARC/IACR MP rules define 8 different histological groups for LC [11]. In contrast, North American cancer registries follow the rules of the Surveillance, Epidemiology, and End Results Program (SEER) [12], considering timing, laterality, and more detailed histology with 34 different groups. Consequently, a second lung adenocarcinoma in a person fully recovered from a previous lung adenocarcinoma diagnosis would not be classified as an SPLC according to IARC/IACR rules, even if the 2 events are 10 years apart. In contrast, SEER would classify the diagnosis as an SPLC because any tumor is considered a new primary LC after three disease-free years. Additionally, even if the disease-free period since the initial diagnosis is shorter, an SPLC would be registered in the US registries if the second adenocarcinoma had an ICD-O-3 histology code that was different at the third number (e.g., bronchiolo-alveolar adenocarcinoma 8250 and papillary adenocarcinoma 8260). Additional file 1: Table S1 shows a detailed comparison of IARC/IACR and SEER MP rules.

Estimating risk for same-site SPC has been an unsolved methodological issue for many years, and experts have warned that using SIR can be misleading [13]. Cross-country comparison is difficult due to varying registration practices [14]. Consequently, recent high-quality publications have not presented respective estimates [4, 15] or highlighted that low-risk estimates contradict increased exposure and may be an artifact of calculating SIR as a relative risk estimate [7]. For cancer registries using IARC/IACR MP rules, SIRs might be underestimated due to the histology-dependent documentation of SPC with the same location as the first cancer. We want to narrow this methodological gap in cancer epidemiology by proposing an adapted risk estimate for same-site SPC that considers the histologic subtype of the first cancer.

Objective

This study aims to assess whether using histology-specific reference rates for estimating the expected number of cases improves the estimation of SIR for SPLC in LC survivors. Therefore, this analysis (i) demonstrates how the conventional SIR metric underestimates the risk for SPLC in LC survivors documented in cancer registries that use IARC/IACR MP rules and (ii) proposes a novel method to calculate histology-specific SIRs to reduce bias and presents updated risk estimates for SPLC in Germany. Finally, we (iii) validate our results by comparing how this new method performs when estimating risk for SPLC in the USA, where multiple primaries are registered following SEER rules.

Methods
Data

This study uses pooled data from 11 regional German population-based cancer registries (PBCRs) that follow IARC/IACR MP rules, covering ~50% (40 million) of the German population [16]. The analysis dataset included all LC survivors diagnosed between 2002 and 2013 who survived at least 6 months after the initial diagnosis. The data source has been previously described [7, 17]; details on case selection and data modifications are described in Additional file 1: Table S2 and Table S3.

As a validation dataset, we pooled data from 17 SEER regions that started registration before 2002 [18]. Details on the included regions and data quality can be found in Additional file 1: Table S4. Methodological aspects of the SEER database were described by Curtis et al. [19].

We set the end of follow-up in both data sources to 31 December 2014 to ensure comparability. We counted SPLC (ICD-10 code C34) and excluded unusual morphology codes for LC that were likely miscoded metastases (codes 8263, 8290, 8720, 8815, 8933, 9050, and 9133).

Our definition of histologically different groups is based on the International Classification of Diseases for Oncology, third edition (ICD-O-3), revision 1, published in 2013 [11]. The four-digit morphology codes were converted into histology groups as listed in Additional file 1: Table S5. To present aggregated results, we additionally

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 3 of 12

grouped the histologic subtypes of LC into squamous cell carcinoma (SCC), adenocarcinoma (AC), small-cell carcinoma (SCLC), large-cell carcinoma (LCC), and other and unspecified (O&U) as suggested by the WHO classification of thoracic tumors [20].

Statistical analysis

Estimating the size of bias for using standard SIR

We simulated various scenarios to estimate the size of bias introduced by using general population reference rates for calculating SIR (ratio of observed to expected cancers O/E) of same-site SPC when IARC/IACR MP rules are applied. This bias occurs because the count for O excludes same-histology SPLC, while the reference rates used to calculate E include same-histology. Our first simulation setting is that the baseline risk of LC survivors to develop an SPLC is the same as that for the general population ($\text{SIR}_{\text{real}} = 1.0$). We determined the proportions of histologically different LC groups p_{hist_j} in the analysis dataset for all index LC cases stratified by sex and assumed that the SPLC would have the same histology group distribution as for the first cancer. To obtain the simulated $\text{SIR}_{\text{simIARC}}$, we multiplied each stratum of obtained expected numbers with a sex- and histology-group correction factor $x_{\text{hist}_j} = 1 - p_{\text{hist}_j}$ for combinations of LC and SPLC that are not possible in our observed cases. In addition to the scenario of no risk difference, we also simulate the impact of this bias when SPLC risk for LC survivors is double that of the general population ($\text{SIR}_{\text{real}} = 2.0$) and a risk increase equivalent to data published by Thakur et al. for US male ($\text{SIR}_{\text{real}} = 3.38$) and female ($\text{SIR}_{\text{real}} = 4.85$) LC survivors [9]. Further details on the simulations can be found in Additional file 1: Section S6.

Calculating a histology-specific SIR

To calculate the conventional SIR (ratio of observed to expected SPLC cases), referred to as SIR1_{raw}, we used age-, sex-, region-, and period-specific general reference rates obtained from the full sample of German PBCR data, including cases with a diagnosis based on a death certificate only (DCO).

Next, we propose a novel method to stratify SIR calculations by histology group of LC (e.g., adenocarcinomas) and use group-specific reference rates excluding same-histology SPLC. The concepts of the conventional SIR1_{raw} and the newly proposed histological subtype-specific SIR2_{sub} are explained in Fig. 1. Most importantly, calculating histological subtype-specific SIR removes the discrepancy in the conventional method of excluding same-histology group tumors from the observed count (O)—due to IARC/IACR MP rules—while including those same-histology tumors in

the reference rates, i.e., expected cases. When calculating the SIR2_{sub}, we always stratified by the histology group of the first tumor. Then, we used histology-specific reference rates IR_{ij} that excluded the same histology group from the incidence. In settings where same-histology SPCs occur in the data (i.e., PBCR not applying IARC/IACR MP rules), those cancers are not counted as observed.

Finally, for all SIR, we aggregated O and E for the total available follow-up from 6 months to a maximum of 13 years and always reported sex-specific estimates. We calculated 95% confidence intervals (CIs) as described by Breslow and Day [21].

Validation analysis using SEER data

To validate our newly proposed method, we apply the SIR2_{sub} measure to cancer registry data that does not follow IARC/IACR MP rules and assess whether this introduces unintended bias. We present descriptive statistics on the validation dataset and interpret comparability with the analysis dataset. Then, we calculate the unadjusted SIR1_{raw} for the USA and SIR2_{sub} using the same method described above. Furthermore, we obtained individual-level information on which cancer cases in the SEER data fulfill the IARC/IACR MP rules (variable INTPRIM). We recalculated the unadjusted SIR for SPLC counted under IARC/IACR rules ($\text{SIR3}_{\text{IARC}}$) and histology-specific SIR ($\text{SIR4}_{\text{subIARC}}$) using patient-level data to receive updated reference rates. To assess the validity of our histology-specific SIR method, we compare the following:

a. SIR2_{sub} and $\text{SIR4}_{\text{subIARC}}$ in the validation dataset with the expectation that they are very similar and lie between SIR1_{raw} and $\text{SIR3}_{\text{IARC}}$ (this would show that the new method delivers a robust estimate, independent of MP rules applied)
b. SIR2_{sub} estimates between Germany and the USA with the expectation that they should be closer than SIR1_{raw} estimates (this is relevant for a scenario where a registry's abidance by IARC/IACR rules is unknown and we want to receive an estimate suitable for cross-country comparison)
c. SIR2_{sub} estimates from Germany to $\text{SIR4}_{\text{subIARC}}$ estimates from the USA to determine the size of residual cross-country differences

Sensitivity analyses

For sensitivity analyses, we restricted our dataset to six German PBCRs with a low DCO rate (smaller than 10%)—an indicator for higher registration quality—and assessed whether this influences the overall risk estimate for SPLC in Germany. Furthermore, we compared the

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 4 of 12

$$SIR = \frac{\sum_{i=1}^{I} count_i}{\sum_{i=1}^{I} IR_i \times pyears_i} = \frac{O}{E}$$

- i stratified by age, sex, region, period, (race)
- IR population reference rates calculated by counting all relevant LC cases in entire registry divided by official population data
- $pyears$ person-years at risk from 6 months after LC until end of follow-up DEC-2014
- O Observed cases of any SPLC
- E Expected cases

- Example:
  Male from Hamburg 54yrs at LC diagnosis in 2007 with Adenocarcinoma survived until end of FU

SIR1$_{raw}$ unadjusted SIR using the conventional method

- i summing up all LC
- IR general population reference rates for LC
- O exclude, but E include same histology LC
- Example: person-years at risk multiplied with **IR of any LC** in males 50-54 years, 2005-2009 from Hamburg
- **SIR3$_{IARC}$** using the conventional method and only counting international primaries

SIR2$_{sub}$ new method using histology-specific reference rates

- i stratifying additionally by histology group of LC j
- IR_{ij} histology-specific reference rates for LC of different subtype
- O and E consistently excluding same histology LC
- Example: person-years at risk multiplied with **IR of non-Adenocarcinoma LC** in males 50-54 years, 2005-2009 from Hamburg
- **SIR4$_{subIARC}$** applying the new method of histology-specific SIR and only counting international primaries

Fig. 1 Comparison of conventional method to calculate standardized incidence ratio (SIR1$_{raw}$) and newly proposed method to stratify by histology group of LC and use group-specific reference rates excluding same histology (SIR2$_{sub}$). The top gray dotted box shows the general formula for SIR, explains the operationalization for each variable in our study setting, and presents an example assessing the risk of SPLC for a survivor with prior adenocarcinoma LC. Icons are visualizing this example. The middle green dotted box explains that in the conventional SIR method, the sample patient is assessed together (unstratified) with the overall risk of other primary LC histologies, and the risk is assessed including all SPLC histologies in the reference rates. In contrast, the bottom violet dotted box shows that the newly proposed method additionally stratifies risk calculation by LC histology, and only different histology SPLCs are counted when calculating the reference rates. *Notes*: Gray arrows mark the option to either not get an SPLC (healthy lung depicted uncolored with a check mark) or be diagnosed with an SPLC of different histology (i.e., different color). Blue, green, and yellow symbolize different histology groups (for simplicity, we only show three). Colored unfilled lungs depict the histology group of the first LC; colored filled lungs depict the histology group of SPLC

results from the German analysis dataset to SEER data for the White population only to account for the majority Caucasian population in Germany.

Software used
All analyses were conducted in R version 4.3.2 [22] with data management using the *tidyverse* and *tidytable*

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 5 of 12

packages [23, 24]. Figures were created with *ggplot2* [25] and tables with *gt* [26]. The routines to calculate reference rates and stratified SIRs are publicly available in the R package *msSPChelpR* [27], and all scripts to create our analyses are published online [28].

Results

Description of the study population
Our analysis dataset comprised 135,572 German LC patients (31.8% females and 68.2% males) who survived for at least 6 months after LC diagnosis and fulfilled our inclusion criteria. The median age at initial diagnosis was 65.6 years in women and 67.2 years in men. Women were most frequently diagnosed with adenocarcinomas (44.8%) and small cell carcinomas (19.8%), while 34.5% of men had squamous cell carcinomas and 31.0% had adenocarcinomas. Patients had a mean follow-up of approximately 2.5 years. Over 70% of patients died after LC without developing another cancer, and another 24% were alive without an SPC event at the end of follow-up. In total, 154 women and 388 men (0.4% of patients) developed an SPLC. This corresponds to the crude incidence rate of 131.1 per 100,000 person-years in women and 168.4 in men. The other 4753 LC survivors (3.5%) developed a second cancer in a location other than the lung and bronchus (Table 1).

Evidence for bias using the conventional SIR estimates for same-site SPC
The most frequent histology groups of primary LC in our analysis dataset were adenocarcinomas (40.8% in women, 28.7% in men), squamous cell carcinomas (16.0% in women, 34.5% in men), and other specific carcinomas (34.1% in women, 27.9% in men) (Additional file 1: Fig. S7). Unspecified carcinomas and unspecified types of cancer comprised 8.8% of LC and, by IARC/IACR rules, could not result in the registration of an SPLC, even if one was diagnosed. In line with IARC/IACR registration rules, none diagnosed SPLC had the same histology group as the index LC in the German data. In contrast, the US validation data showed that the SEER registries had between 27.3 and 54.2% same histology group SPLC (Additional file 1: Table S8).

The simulation shows that real SIRs are underestimated when IARC/IACR rules are applied for SPLC. In the hypothetical case of $SIR_{real} = 1.0$, the simulated $SIR_{simIARC}$ is 0.71 (95% CI 0.52–0.93) for women and 0.73 (95% CI 0.65–0.81) for men, incorrectly suggesting a risk reduction instead of no risk difference. A doubling of the risk in real SIR would translate to a $SIR_{simIARC}$ of 1.41 for women and 1.45 for men. When simulating the US rates under IARC/IACR MP rules, we obtain a $SIR_{simIARC}$ of 2.46 instead of $SIR_{real} = 3.38$ for men and a $SIR_{simIARC}$ of

3.42 instead of $SIR_{real} = 4.85$ for women. Despite the differences in predominant histology groups of LC between men and women, we observed a similar underestimation of real SIR by approximately 30% for both sexes (Table 2).

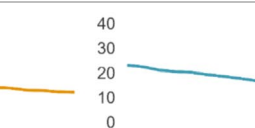### Risk estimates for SPLC after LC in Germany using histology-specific SIR
The adjusted relative risk for LC survivors in Germany to develop an SPLC was 2.98 (95% CI 2.53–3.49) for females and 1.15 (95% CI 1.03–1.27) for males. The histological subtypes of index LC with the highest relative risk for SPLC were squamous cell carcinoma ($SIR2_{sub} = 5.17$, 95% CI 3.94–6.67) in women and SCLC ($SIR2_{sub} = 1.36$, 95% CI 0.99–1.81) in men. Figure 2 shows that the proposed method of using histological subtype-specific SIR resulted in higher risk estimates. This means that under this adjusted risk estimate, male LC survivors have a 15% higher risk for SPLC than the German male population. A risk increase was observed for all histological subtypes except for large-cell carcinoma. In contrast, the unadjusted SIR resulted in a 15% lower risk than in the male population ($SIR1_{raw} = 0.85$, 95% CI 0.77–0.94). For female LC survivors, the new method further increased risk estimates, particularly for adenocarcinoma ($SIR2_{sub} = 2.53$, 95% CI 1.91–3.28 vs. $SIR1_{raw} = 1.69$, 95% CI 1.28–2.19). When stratifying the analysis by time since LC diagnosis, the $SIR2_{sub}$ increased with more prolonged survival for both sexes (Additional file 1: Fig. S9).

In the sensitivity analysis based on 75,272 person-years at risk from German PBCRs with low DCO rates, 106 SPLC occurred (19.6% of the initial analysis dataset). Both $SIR1_{raw}$ and $SIR2_{sub}$ were considerably lower than in the main analysis ($SIR2_{sub} = 2.30$ in females, $SIR2_{sub} = 0.99$ in males), but the overall pattern of substantial risk estimate increase using $SIR2_{sub}$ remained (Additional file 1: Table S10).

Validation analysis using SEER data
Table 1 shows that our analysis dataset (German ZfKD data from 11 regions) and the validation dataset (US SEER data from 17 regions) are comparable concerning most characteristics of included LC survivors with similar age structure, histology group distribution, and patient status. However, Germany still has a growing age-standardized incidence rate for women, and the rate for men is at a higher level than in the USA. We also see that crude incidence rates for SPC in locations other than the lung and bronchus are similar for Germany (females 1137.9 and males 1483.0 per 100,000 person-years) and the USA (females 1099.4 and males 1489.6 per 100,000 person-years). The incidence of SPLC, however, differs greatly—the USA is showing 6 to 8 times higher crude

Table 1 Characteristics of the analyzed study population with primary lung cancer. Age-standardized incidence rates of primary lung cancer (ASIR), follow-up time, characteristics of patients included in the main analysis with at least 6 months of survival, and absolute incidence of second primary cancer (SPC) by sex

	Analysis Dataset – Germany (ZfKD data from 11 regions)		Validation Dataset -- United States (SEER data from 17 regions)	
	Female	Male	Female	Male
Age-standardized incidence rate of lung cancer				
ASIR 2002 - 2013				
ASIR in 2013 [per 100,000] (95% CI)	16.4 (16.0 to 16.9)	34.1 (33.5 to 34.7)	12.0 (11.9 to 12.1)	14.9 (14.8 to 15.0)
Patients with primary LC (with at least 6 months of survival)[a]				
n (% of Total)	43,175 (31.8%)	92,397 (68.2%)	133,401 (50.6%)	130,421 (49.4%)
Age at diagnosis of LC				
30 - 49	4,065 (9.4%)	5,373 (5.8%)	8,653 (6.5%)	7,496 (5.7%)
50 - 59	9,990 (23.1%)	17,717 (19.2%)	23,306 (17.5%)	24,390 (18.7%)
60 - 69	13,996 (32.4%)	33,674 (36.4%)	39,836 (29.9%)	42,210 (32.4%)
70 - 79	11,380 (26.4%)	28,998 (31.4%)	40,486 (30.3%)	39,071 (30.0%)
80+	3,744 (8.7%)	6,635 (7.2%)	21,120 (15.8%)	17,254 (13.2%)
Median age [years]	65.6	67.2	68.5	68.5
Year of diagnosis of LC				
2002 - 2005	11,931 (27.6%)	30,235 (32.7%)	43,382 (32.5%)	44,697 (34.3%)
2006 - 2009	14,712 (34.1%)	31,499 (34.1%)	44,922 (33.7%)	43,957 (33.7%)
2010 - 2013	16,532 (38.3%)	30,663 (33.2%)	45,097 (33.8%)	41,767 (32.0%)
Histology of LC				
Squamous cell carcinoma (SCC)	6,886 (15.9%)	31,858 (34.5%)	21,280 (16.0%)	33,214 (25.5%)
Adenocarcinoma (AC)	19,327 (44.8%)	28,626 (31.0%)	59,518 (44.6%)	47,065 (36.1%)
Small cell carcinoma (SCLC)	8,530 (19.8%)	15,795 (17.1%)	17,307 (13.0%)	15,717 (12.1%)
Large cell carcinoma (LCC)	2,616 (6.1%)	5,652 (6.1%)	7,343 (5.5%)	7,667 (5.9%)
Other & unspecified (O&U)	5,816 (13.5%)	10,466 (11.3%)	27,953 (21.0%)	26,758 (20.5%)
Person-years at risk				
Mean follow-up [months]	32.7	29.9	34.3	29.9
Sum of PYAR	117496	230346	380943	324648
Patient status				
SPLC developed	154 (0.4%)	388 (0.4%)	3,775 (2.8%)	3,102 (2.4%)
other SPC developed	1,337 (3.1%)	3,416 (3.7%)	4,188 (3.1%)	4,836 (3.7%)
dead after LC	29,300 (67.9%)	68,582 (74.2%)	90,026 (67.5%)	96,168 (73.7%)
no event until end of follow-up	12,384 (28.7%)	20,011 (21.7%)	35,412 (26.5%)	26,315 (20.2%)
Absolute incidence rate of SPC				
SPLC IR [per 100,000 PYAR] (95% CI)	131.1 (111.2 to 153.5)	168.4 (152.1 to 186.1)	991.0 (959.6 to 1023.1)	955.5 (922.2 to 989.7)
Other SPC IR [per 100,000 PYAR] (95% CI)	1137.9 (1077.7 to 1200.6)	1483.0 (1433.7 to 1533.6)	1099.4 (1066.3 to 1133.2)	1489.6 (1447.9 to 1532.2)

ASIR Age-standardized incidence rate based on the World Standard Population 1960, *DCO* Death-certificate only, *IR* Incidence rate, *LC* Primary lung cancer, *PYAR* Person-years at risk, *SPC* Second primary cancer, *SPLC* Second primary lung cancer

[a] After exclusion by age and unusual histology

rates than Germany (991.0 vs. 131.1 per 100,000 person-years for women and 955.5 vs. 168.4 for men).

The unadjusted risk estimates for SPLC in US LC survivors under SEER MP rules were more than twofold the rates under IARC/IACR MP rules ($SIR1_{raw} = 5.52$ and $SIR3_{IARC} = 2.52$ in women; $SIR1_{raw} = 3.77$ and $SIR3_{IARC} = 1.71$ in men). When using the histology-specific method, the risk estimates converge for overall SPLC risk ($SIR2_{sub} = 4.37$ and $SIR4_{subIARC} = 3.58$ in women;

$SIR2_{sub} = 2.94$ and $SIR4_{subIARC} = 2.34$ in men) and for most subgroups presented in Table 3.

Comparing estimates between Germany and the USA, we observed that $SIR2_{sub}$ estimates are much closer than $SIR1_{raw}$. In women, the gap narrows from 2.14 vs. 5.52 for unadjusted SIR to 2.98 vs. 4.37. In men, the wide unadjusted gap of 0.85 vs. 3.77 narrows slightly to 1.15 vs. 2.94. When examining the histology-specific estimate under IARC/IACR MP rules ($SIR4_{subIARC}$), we still see a

Table 2 Estimating the risk for SPLC under IARC/IACR multiple primary rules SIR$_{simIARC}$ given an assumed true risk SIR$_{real}$

|	SIR$_{real}$ = 1.0 (No effect)		SIR$_{real}$ = 2.0 (Double the risk)		SIR$_{real}$ = 3.38 (SEER males)[1]		SIR$_{real}$ = 4.85 (SEER females)[2]	
	SIR$_{simIARC}$	SIR$_{real}$	SIR$_{simIARC}$	SIR$_{real}$	SIR$_{simIARC}$	SIR$_{real}$	SIR$_{simIARC}$	SIR$_{real}$
Female	**0.71**	1.00	**1.41**	2.00			**3.42**	4.85
Male	**0.73**	1.00	**1.45**	2.00	**2.46**	3.38		

[1] SIR for SPLC after LC for Males in SEER according to Thakur et al. 2018

[2] SIR for SPLC after LC for Females in SEER according to Thakur et al. 2018

SIR$_{simIARC}$ simulated SIR under IARC/IACR multiple primary rules, not allowing same-histology SPLC; SIR standardized incidence ratio; PYAR person-years at risk; SPC second primary cancer; SPLC second primary lung cancer

Sex-specific relative risk for SPLC in German lung cancer survivors

Estimation of risk for SPLC after LC using general reference rates (SIR1$_{raw}$) and subtype-specific reference rates excluding same-histology group (SIR2$_{sub}$)
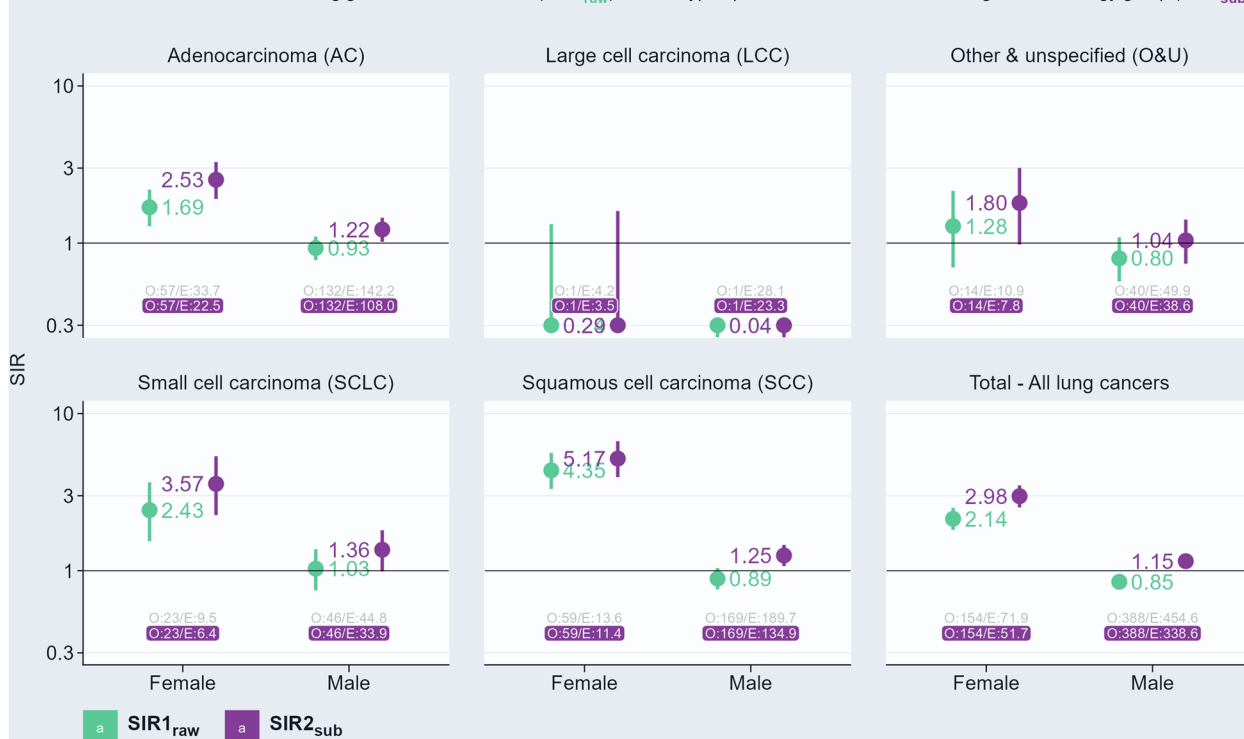
Fig. 2 Sex-specific relative risk for SPLC in German lung cancer survivors. Estimation of risk for SPLC after LC using general reference rates (SIR1$_{raw}$) shown in green and subtype-specific reference rates excluding same-histology group (SIR2$_{sub}$) shown in purple. SIR estimates are presented stratified by histologic subtype (using the WHO classification of thoracic tumors) as well as a total of all histologies. Within each facet, SIR for females is shown on the left and for males on the right with the length of the line indicating the respective 95% CI. Observed and expected counts below in gray belong to the conventional method (SIR1$_{raw}$) and those highlighted in purple yield the newly proposed risk estimate (SIR2$_{sub}$)

residual country difference with only slightly lower SPLC risk in German women (2.98 vs. 3.58) but half the risk for males compared to the USA (1.15 vs. 2.34). For some subgroups with a small number of observed cases in the German data, such as patients with large-cell carcinoma or below age 50, the risk gap between analysis and validation data may remain high, even after using the novel SIR estimate (Table 3).

The sensitivity analyses, comparing Germany to the White population in SEER, did not alter the results,

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 8 of 12

Table 3 Validation analysis—risk for SPLC using unadjusted and histology-specific SIR method

	Germany (Analysis dataset - IARC/IACR MP rules)				United States (Validation dataset - SEER MP rules)						
	$SIR1_{raw}$	$SIR2_{sub}$	95% CI_{SIR2}	O_{SIR2}	$SIR1_{raw}$	$SIR2_{sub}$	95% CI_{SIR2}	$SIR3_{IARC}$	$SIR4_{subIARC}$	O_{SIR1}	O_{SIR2}
Females											
Total - All lung cancers	2.14	**2.98**	2.53–3.49	154	5.52	**4.37**	4.18–4.56	2.52	3.58	3,775	2,106
Histology of LC											
Adenocarcinoma (AC)	1.69	**2.53**	1.91–3.28	57	6.08	**4.48**	4.20–4.76	2.61	3.96	2,055	996
Large cell carcinoma (LCC)	0.24	**0.29**	0.01–1.60	x	4.04	**3.95**	3.27–4.73	0.32	0.36	139	119
Other & unspecified (O&U)	1.28	**1.80**	0.98–3.01	14	4.04	**3.88**	3.49–4.30	2.07	3.00	551	364
Small cell carcinoma (SCLC)	2.43	**3.57**	2.26–5.35	23	4.26	**4.51**	3.79–5.32	2.52	3.96	206	139
Squamous cell carcinoma (SCC)	4.35	**5.17**	3.94–6.67	59	6.50	**4.66**	4.26–5.09	3.37	4.08	824	488
Age at diagnosis of LC											
30 - 49	4.20	**6.39**	2.57–13.17	7	37.95	**32.88**	26.97–39.70	15.58	24.76	198	108
50 - 59	3.56	**5.22**	3.79–7.00	44	14.29	**12.06**	10.85–13.37	6.65	10.18	656	362
60 - 69	2.37	**3.35**	2.59–4.26	66	7.23	**5.94**	5.54–6.36	3.36	4.91	1,456	819
70 - 79	1.34	**1.81**	1.24–2.55	32	3.90	**3.06**	2.83–3.30	1.75	2.46	1,224	681
80+	0.82	**1.06**	0.34–2.47	5	2.05	**1.55**	1.30–1.83	0.92	1.23	241	136
Year of diagnosis of LC											
2002 - 2005	2.27	**3.07**	2.31–3.99	55	5.71	**4.44**	4.15–4.74	2.62	3.66	1,620	900
2006 - 2009	1.91	**2.66**	1.99–3.48	53	5.61	**4.48**	4.18–4.80	2.60	3.70	1,452	817
2010 - 2013	2.31	**3.33**	2.44–4.44	46	5.00	**4.02**	3.63–4.44	2.17	3.15	703	389
Males											
Total - All lung cancers	0.85	**1.15**	1.03–1.27	388	3.77	**2.94**	2.81–3.08	1.71	2.34	3,102	1,773
Histology of LC											
Adenocarcinoma (AC)	0.93	**1.22**	1.02–1.45	132	4.19	**3.13**	2.91–3.37	1.98	2.81	1,344	707
Large cell carcinoma (LCC)	0.04	**0.04**	0.00–0.24	x	2.79	**2.92**	2.42–3.48	0.25	0.28	135	123
Other & unspecified (O&U)	0.80	**1.04**	0.74–1.41	40	2.68	**2.58**	2.29–2.89	1.38	1.91	415	289
Small cell carcinoma (SCLC)	1.03	**1.36**	0.99–1.81	46	3.09	**3.38**	2.80–4.04	2.13	3.12	160	120
Squamous cell carcinoma (SCC)	0.89	**1.25**	1.07–1.46	169	4.22	**2.86**	2.62–3.11	1.77	2.35	1,048	534
Age at diagnosis of LC											
30 - 49	1.64	**2.24**	0.73–5.22	5	26.40	**22.55**	17.33–28.86	13.37	19.21	106	63
50 - 59	1.61	**2.20**	1.70–2.81	65	10.10	**8.01**	7.14–8.95	4.35	6.12	547	308
60 - 69	1.16	**1.57**	1.37–1.80	207	4.79	**3.74**	3.46–4.03	2.17	3.01	1,199	673
70 - 79	0.53	**0.71**	0.58–0.86	104	2.73	**2.16**	1.99–2.35	1.28	1.75	1,028	596
80+	0.20	**0.25**	0.10–0.51	7	1.60	**1.26**	1.06–1.49	0.69	0.91	222	133
Year of diagnosis of LC											
2002 - 2005	0.79	**1.05**	0.89–1.22	156	3.71	**2.92**	2.72–3.13	1.72	2.32	1,352	785
2006 - 2009	0.94	**1.26**	1.07–1.48	157	4.07	**3.15**	2.92–3.40	1.82	2.49	1,239	702
2010 - 2013	0.84	**1.15**	0.90–1.44	75	3.30	**2.59**	2.30–2.91	1.50	2.10	511	286

Notes: O_{SIR1} number of cases observed in the data for $SIR1_{raw}$; O_{SIR2} number of cases observed in the data for $SIR2_{sub}$, ZfKD data $O_{SIR1} = O_{SIR2}$; SEER Surveillance, Epidemiology, and End Results Program; SIR standardized incidence ratio; $SIR1_{raw}$ unadjusted SIR using age-, sex-, region-, period-specific reference rates; $SIR2_{sub}$ histological subtype-specific SIR using age-, sex-, region-, period- and histology-specific reference rates and excluding same-histology group SPLC from observed and expected; $SIR3_{IARC}$ unadjusted SIR but only counting international primaries (IARC/IACR MP rules), for ZfKD data $SIR1_{raw} = SIR3_{IARC}$; $SIR4_{subIARC}$ histological subtype-specific SIR but only counting international primaries (IARC/IACR MP rules), for ZfKD data $SIR2_{sub} = SIR4_{subIARC}$; SPLC second primary lung cancer; x censored counts of observed smaller than 5 for data privacy reasons; ZfKD German Centre for Cancer Registry Data

showing that previous stratification of SIR by race adequately accounts for risk differences between races (Additional file 1: Table S11).

Discussion

In this epidemiological study on the risk of SPLC, we demonstrated a bias introduced by the standard approach of calculating SIR for same-site SPC in cancer registries that use IARC/IACR MP rules. Our simulations showed an underestimation of true SIR by approximately 30% in both sexes because general population reference rates include tumors that cannot be observed in the at-risk population.

To minimize this bias, we proposed the method of histology-specific SIR and presented adjusted numbers to estimate the relative risk of SPLC in LC survivors in Germany. Using this method, the risk of being diagnosed with an SPLC was 2.98-fold for female LC survivors and

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 9 of 12

1.15-fold for male survivors compared to the general population. These numbers are substantially higher than previously published unadjusted SIR estimates using the standard approach (i.e., 2.06 and 0.83 for female and male LC survivors, respectively) [7].

Our validation analysis showed that the risk estimates for SPLC under SEER MP rules and the risk under IARC/IACR MP rules converge when using the proposed histology-specific compared to the standard approach. Moreover, the SPLC risk in Germany and the USA became comparable under the new method. In the descriptive analysis, we observed a similar incidence of primary LC, but an eight times higher crude incidence of SPLC in the USA compared to Germany, due to different registration rules and population structure. We narrowed this gap to an about twofold increased risk in the most comparable setting of histology-specific SIR counting international primaries only. Thus, we showed that the proposed histology-specific method makes the SIR metric more robust against differing or unclear MP rules and, therefore, more suitable for cross-country comparisons.

With the availability of long-term cancer registry data from SEER, relative risk estimates and survival analyses were predominantly published for the USA. Thakur et al. used 13 SEER registries for the period 1992 to 2007 and found a SIR estimate of 4.85 (4.66–5.05) in women and 3.38 (3.24–3.52) in men [9]. Our $SIR1_{raw}$ estimates for a later period with more included regions were slightly higher for both sexes but—as expected for the same data source—showed very similar risk increases for younger age and adenocarcinoma index LC.

Other national population-based estimates of SPLC risk among LC survivors are sparse but include a study by Barclay et al. from the England cancer registry that showed an overall SIR for SPLC of 1.47 (not stratified by sex) for the same period, also using IARC/IACR MP rules and very similar case inclusion as our study [8]. They included almost 30% of unspecified LC diagnoses, which might explain their low overall unadjusted SIR estimate.

For Germany, we observed that using histology-specific calculation of risk for SPLC resulted in very different conclusions than the standard method. The previously published risk reduction for male LC survivors [7] turns into a risk increase in our updated analysis. Similarly, for the most frequent subtypes of LC, the risk of SPLC increases from a protective or null effect to a significant risk increase for LC survivors compared to the general population. In line with previous findings from Germany, the USA, and England, females have a higher relative risk for SPLC than males [7–9]. The results of bias simulation apply to other PBCRs using IARC/IACR MP rules because they depend only on the distribution of histology groups, which are similar across countries [29].

The SIR is an established measure in cancer epidemiology and beyond to identify at-risk populations and test etiologic hypotheses [19]. However, our analysis revealed the problem that the SIR metric can be seriously biased when observed and expected counts are based on different logics. Other examples are handling the intensified screening period of cancer survivors shortly after diagnosis [19], including DCO cases in the expected count [30], or stratifying by race.

A critical discussion in the analysis of SPC risk is the question of what constitutes biologically independent cancers. Although there might be a pathological answer, strict (IARC) and less stringent (SEER) rules for registering multiple primary cancers indicate different approaches. While IARC/IACR MP rules were historically focused on excluding potential metastases or recurrences, they give less importance to shared risk factors such as smoking, which can lead to preventable new cancers with the same appearance as the previous disease. Therefore, the SEER approach of registering SPC independent of histological differences and instead taking time into account has its merits. From the patient's perspective, it is less relevant if, after a phase of complete remission, a new cancer of the same subtype occurs in the same site or whether this is an unexpected late recurrence. Both necessitate acute treatment according to the most recent guidelines and come with substantial psychological stress. From an epidemiological perspective, the differentiation between what constitutes a true SPC and late recurrence can be seen as arbitrary. Both are relevant to evaluating the disease risk of cancer survivors and play a role in the discussion about the need for increased surveillance. In particular, contralateral lung cancers are often underreported and instead seen as the spread of the original cancer [31]. On the other hand, it has been shown that certain multifocal tumors in some organs are likely to originate from the same transformed cells [32] and thus would not have been preventable. A potential solution to this dilemma could be to enrich existing PBCR with more information on metastases and recurrence and report the burden of both. Furthermore, improving the quality of cancer registries, e.g., increasing the share of microscopically verified cases, would reduce the role of unspecified LC and thus increase the precision of SPC differentiation under IARC/IACR MP rules.

This analysis also highlights the need to improve the clarity of IARC/IACR MP rules that have not been updated since 2004 and do not reflect improvements in (more fine-grained) histological classification since then. A simplified version using the broad LC categories recently published in the WHO Reporting System for Lung Cytopathology could be the common basis for a classification update [20, 33].

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 10 of 12

The importance of methodological choices on the validity of the SIR metric to determine and compare group-specific cancer risk has already been discussed. Crocetti et al. highlighted that we need to be clear about which cases to include in the observed count and the importance of applying the same logic when making cross-country comparisons and comparing extensively screened patients to the general population reference rates [34]. Our research extends this concern and highlights that including expected cases in the denominator of the SIR metric while these cases can never be observed—due to MP rules forbidding their registration—will introduce bias and result in an underestimation of cancer risk.

This study is the first nationwide analysis to provide reliable estimates of SPLC risk in German LC survivors. It adds to the sparse international numbers, as PBCR-based estimates have only been published for the USA and England. We newly developed the method of histological subtype-specific SIR that can reduce bias caused by using general population reference rates in determining the relative risk of same-site SPC for cancer registry data based on IARC/IACR MP rules. The method increases the comparability of SIR across countries when different MP rules are applied, information on registration practices is missing, or registration practices vary over time within a single PBCR. Furthermore, our study is the first to compare the relative risk of SPLC between Germany and the USA and provide estimates that allow SEER data to be used in international comparison.

Our proposed method has several limitations. In the situation in which PBCR data is incomplete, i.e., incident SPCs are not recorded in the registry, the proposed method still underestimates the SIR. This is especially relevant for IARC-compliant registries where SPCs are not recorded if the histology of the index tumor is unspecified. Moreover, the method works best for cancer entities that show a uniform distribution of incidence across histological subtypes. For example, the subtype is not informative in breast cancer, where adenocarcinoma comprises more than 90% of all malignancies. We encourage researchers with access to cancer registry data from other countries to test the proposed method. With regard to the estimation of SPLC risk for Germany, our study results are limited by varying completeness and quality of regional PBCR data—in particular, the existence of tumors without histological confirmation, a high share of DCO diagnoses, and incomplete follow-up of patients in cases of migration or partial record linkage with death registry data. Finally, our validation analysis also showed that for registry data under SEER MP rules, the histology-specific SIR generally suggests a lower risk than the unadjusted SIR (except primary SCLC). This could be an underestimation of risk if the true SIR of same-histology SPLC is higher than for different-histology tumors.

Conclusions

In this study, we newly developed a method of histological subtype-specific SIR that can reduce bias when estimating the relative risk of same-site SPC. Our study shows that estimations for the relative risk of SPLC after LC are heavily influenced by differing registration practices for multiple primaries and how data is analyzed. Our approach of using histology-specific reference rates improves SIR estimations and facilitates cross-country comparison, especially for PBCRs that apply IARC/IACR MP rules. Future research can use this method to update epidemiological data, allowing clinicians to make informed decisions about patients' SPLC risk based on accurate temporal and regional incidence trends. Differing registration practices remain an obstacle when comparing relative risk for SPC across countries and over time. Therefore, it should be further evaluated how MP rules and methods to determine SPC risk in the same location can be harmonized.

Abbreviations

CI	Confidence interval
E	Expected cases
IACR	International Association for Cancer Registration
IARC	International Agency for Research on Cancer
LC	Lung cancer
MP	Multiple primaries
O	Observed cases
PBCR	Population-based cancer registry
SEER	Surveillance, Epidemiology, and End Results Program
SIR	Standardized incidence ratio
$SIR1_{raw}$	Unadjusted SIR using the conventional method
$SIR2_{sub}$	Proposed new method using histology-specific reference rates to calculate SIR
$SIR3_{IARC}$	Unadjusted SIR using the conventional method and only counting international primaries for both observed and expected cases
$SIR4_{subIARC}$	Applying the new method of histology-specific SIR and only counting international primaries
SIR_{real}	Assumed true SIR in reality
$SIR_{simIARC}$	Simulated SIR when using international multiple primary rules
SPLC	Second primary lung cancer
ZfKD	German Centre for Cancer Registry Data

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-024-03398-9.

Additional file 1: Table S1. Comparison of IARC/IACR and SEER multiple primary rules. **Table S2.** Details of dataset filtering. **Table S3.** Details of data modifications. **Table S4.** Data quality for included regions and SIR estimates. **Table S5.** Conversion table of histology codes into ICD-O-3 histologically 'different' groups and histological subtypes of lung cancer. **Section S6.** Details on simulations to estimate the size of bias using standard SIR. **Fig. S7.** Histological groups of LC and SPLC. **Table S8.** Frequency of same-histology SPLC by region. **Fig. S9.** Relative risk for SPLC in lung cancer survivors stratified by follow-up time. **Table S10.** Sensitivity analysis

---

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 11 of 12

A – Risk of SPLC using unadjusted and histology-specific SIR method [restricted to six German PBCR with low DCO rate]. **Table S11.** Sensitivity analysis B – Risk of SPLC using unadjusted and histology-specific SIR method [SEER restricted to White population].

Availability of data and materials

Due to legal restrictions, the individual-level raw data used for this analysis (i.e., German cancer registry data) is only available via request to the German Centre for Cancer Registry Data (ZfKD), which can provide a scientific use file. More information on the application process is provided on the ZfKD website (https://www.krebsdaten.de/Krebs/EN/Content/ScientificUseFile/scientific usefile_node.html).
The validation dataset (i.e., US cancer registry data) is publicly available via the Surveillance, Epidemiology, and End Results Program (SEER). More information on the data request process is provided on the SEER website (https://seer. cancer.gov/data/access.html).
The authors of this paper strongly support open science and have therefore published both the analysis code and the newly programmed software to review and download under a GPL-3 license. The respective references are provided in the manuscript.
Repository for analysis code: https://github.com/marianschmidt/pub_spc_ sirmethods_bmed
Repository for software code: https://doi.org/https://doi.org/10.5281/zenodo. 5055870.

Declarations

Ethics approval and consent to participate

This study analyzed the scientific use file (SUF) of the German Centre for Cancer Registry Data (ZfKD). Data collection is mandated and regulated by the German Federal Law on Cancer Registration (BKRG), and secondary data analysis of the anonymous SUF—as used in this study—does not require ethics approval or consent by patients (§3 and §5 BKRG).
The validation dataset (i.e., US cancer registry data) is publicly available and de-identified and thus does not require ethics approval in accordance with local laws.

Consent for publication

Not applicable.

Competing interests

ME is employed by Daiichi Sankyo Europe GmbH since July 2023, after the completion of the study.
All other authors declare that they have no competing interests.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Zhou B, Zang R, Zhang M, Song P, Liu L, Bie F, et al. Worldwide burden and epidemiological trends of tracheal, bronchus, and lung cancer: a population-based study. EBioMedicine. 2022;78:103951.
3. Aredo JV, Luo SJ, Gardner RM, Sanyal N, Choi E, Hickey TP, et al. Tobacco smoking and risk of second primary lung cancer. J Thorac Oncol. 2021;16:968–79.
4. Boakye EA, Buchanan P, Hinyard L, Osazuwa-Peters N, Simpson MC, Schootman M, et al. Trends in the risk and burden of second primary malignancy among survivors of smoking-related cancers in the United States. Int J Cancer. 2019;145:143–53.
5. Park ER, Japuntich SJ, Rigotti NA, Traeger L, He Y, Wallace RB, et al. a snapshot of smokers following lung and colorectal cancer diagnosis. Cancer. 2012;118:3153–64.
6. Walker MS, Vidrine DJ, Gritz ER, Larsen RJ, Yan Y, Govindan R, et al. Smoking relapse during the first year after treatment for early-stage non-small-cell lung cancer. Cancer Epidemiol Biomarkers Prev. 2006;15:2370–7.
7. Eberl M, Tanaka LF, Kraywinkel K, Klug SJ. Incidence of smoking-related second primary cancers after lung cancer in Germany: an analysis of nationwide cancer registry data. J Thorac Oncol. 2022;17:388–98.
8. Barclay ME, Lyratzopoulos G, Walter FM, Jefferies S, Peake MD, Rintoul RC. Incidence of second and higher order smoking-related primary cancers following lung cancer: a population-based cohort study. Thorax. 2019;74:466–72.
9. Thakur MK, Ruterbusch JJ, Schwartz AG, Gadgeel SM, Beebe-Dimmer JL, Wozniak AJ. Risk of second lung cancer in patients with previously treated lung cancer: analysis of Surveillance, Epidemiology, and End Results (SEER) data. J Thorac Oncol. 2018;13:46–53.
10. IARC Working Group Report. International rules for multiple primary cancers (ICD-O third edition). Eur J Cancer Prev. 2005;14:307–8.
11. Fritz AG, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al editors. International classification of diseases for oncology: ICD-O (3rd edition, 1st revision). First Revision. Geneva: World Health Organization; 2013.
12. Johnson C, Peace S, Adamo P, Fritz A, Percy-Laurry A, Edwards BK. The 2007 multiple primary and histology coding rules. Bethesda, MD; 2007. https://seer.cancer.gov/tools/mphrules/download.html. Accessed 19 Apr 2024.
13. Jégu J, Moitry M, Bara S, Trétarre B, Guizard A-V, Woronoff A-S, et al. Methodological issues of assessing the risk of a second cancer occurring in the same site as a first cancer using registry data. Cancer Epidemiol. 2017;51:41–3.
14. Thun MJ, Linet MS, Cerhan JR, Haiman C, Schottenfeld D, editors. Cancer epidemiology and prevention. 4th ed. 2017.
15. Sung H, Hyun N, Leach CR, Yabroff KR, Jemal A. Association of first primary cancer with risk of subsequent primary cancer among survivors of adult-onset cancers in the United States. JAMA. 2020;324:2521–35.
16. Zentrum Für Krebsregisterdaten (ZfKD) Im Robert Koch-Institut. Datensatz des ZfKD auf Basis der epidemiologischen Landeskrebsregisterdaten Epi2016_2, verfügbare Diagnosejahre bis 2014. https://www.da-ra.de/dara/study/web_show?res_id=626806&mdlang=en&detail=true . 2017.
17. Arndt V, Holleczek B, Kajüter H, Luttmann S, Nennecke A, Zeissig SR, et al. Data from population-based cancer registration for secondary data analysis: methodological challenges and perspectives. Gesundheitswesen. 2020;82(S 01):S62-71.
18. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat database: Incidence - SEER 9+13+18 regs merged research data + Hurricane Katrina impacted Louisiana cases, Nov 2018 sub (1975–2016 varying) linked to county attributes - time dependent (1990–2016) total U.S., 1969–2017 counties, released April 2019, based on the November 2018 submission. https://www.seer.cancer.gov.
19. Curtis RE, Freedman DM, Ron E, Ries LAG, Hacker DG, Edwards BK, et al., editors. New malignancies among cancer survivors: SEER Cancer Registries, 1973–2000. Bethesda, MD: National Cancer Institute, NIR Publ. No. 05–5302; 2006.
20. World Health Organization, IARC, editors. Thoracic tumours. 5th ed. Lyon: International Agency for Research on Cancer; 2021.

Eberl *et al. BMC Medicine*   (2024) 22:187

Page 12 of 12

21. Breslow NE, Day NE. Statistical methods in cancer research volume II: the design and analysis of cohort studies. Lyon: International Agency for Research on Cancer; 1987.
22. R Core Team. R: a language and environment for statistical computing. 2023.
23. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4:1686.
24. Fairbanks M. Tidytable: tidy interface to 'data.table'. 2020. https://CRAN.R-project.org/package=tidytable. Accessed 19 Apr 2024.
25. Wickham H. Ggplot2: elegant graphics for data analysis. New York: Springer; 2016.
26. Iannone R, Cheng J, Schloerke B. Gt: easily create presentation-ready display tables. 2020. https://CRAN.R-project.org/package=gt. Accessed 19 Apr 2024.
27. Eberl M. msSPChelpR: helper functions for second primary cancer analyses. 2024. https://doi.org/10.5281/zenodo.4084961 . Accessed 19 Apr 2024.
28. Eberl M. Code for Paper 'Histology-specific standardized incidence ratio improves the estimation of second primary lung cancer risk' - marianschmidt/pub_spc_sirmethods_bmed. 2023. https://github.com/marianschmidt/pub_spc_sirmethods_bmed. Accessed 19 Apr 2024.
29. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, et al., editors. Cancer incidence in five continents, Vol. XI (electronic version). Lyon: International Agency for Research on Cancer; 2021. https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Cancer-Incidence-In-Five-Continents%C2%A0Volume-XI-2021. Accessed 19 Apr 2024.
30. Eberl M, Tanaka LF. Effect of counting DCO cases when calculating the relative incidence of second primary cancer (presentation). 2020.
31. Johnson BE. Second lung cancers in patients after treatment for an initial lung cancer. J Natl Cancer Inst. 1998;90:1335–45.
32. Werra C de, Donzelli I, Perone M, Micco RD, Orabona G. Multifocal and multicentric tumors. Renda A, editor. Milano: Springer Milan; 2009. p. 129–42.
33. IAC-IARC-WHO Joint Editorial Board, editor. WHO Reporting System for Lung Cytopathology. 1st ed. Lyon: International Agency for Research on Cancer; 2022.
34. Crocetti E, Buzzoni C, Giuliani O. Methods for second primary cancer evaluation have to be standardized. Int J Cancer. 2018;142:1285–95.

Publisher's Note