THERAPEUTIC ADVANCES in
Medical Oncology

Original Research

Handling missing covariates in observational studies: an illustration with the assessment of prognostic factors of survival outcomes in soft-tissue or visceral sarcomas in irradiated fields (SIF)

Ther Adv Med Oncol

2024, Vol. 16: 1–11

DOI: 10.1177/17588359231220999

© The Author(s), 2024.
Article reuse guidelines: sagepub.com/journals-permissions

Noémie Huchet, Nicolas Penel, Sylvie Bonvalot, Juliette Thariat, Françoise Ducimetière, Antoine Giraud, Maud Toulmonde, Axel Le Cesne, Jean-Yves Blay and Carine Bellera

Abstract

Background: Missing covariates are common in observational research and can lead to bias and loss of statistical power. Limited data regarding prognostic factors of survival outcomes of sarcomas in irradiated fields (SIF) are available. Because of the long lag time between irradiation of first cancer and scarcity of SIF, missing data are a critical issue when analyzing long-term outcomes. We assessed prognostic factors of overall (OS), progression-free (PFS), and metastatic-progression-free (MPFS) survivals in SIF using three methods to account for missing covariates.

Methods: We relied on the NETSARC French Sarcoma Group database, Cox (OS/PFS), and competitive hazards (MPFS) survival models. Covariates investigated were age, sex, histological subtype, tumor size, depth and grade, metastasis, surgery, surgical resection, surgeon's expertise, imaging, and neo-adjuvant treatment. We first applied multiple imputation (MI): observed data were used to estimate the missing covariate. With the missing-data modality approach, a category missing was created for qualitative variables. With the complete-case (CC) approach, analysis was restricted to patients without missing covariates.

Results: CC subjects ($N = 167$; 33%) presented more often with soft-tissue sarcoma (*versus* visceral sarcoma) and grade I–II tumors as compared to the 504 eligible cases. With MI ($N = 504$), factors associated with the worst outcome included metastasis ($p = 0.04$) and R1/R2 resection ($p < 0.001$) for OS; higher grade/non-gradable tumors ($p = 0.002$) and R1/R2 resection ($p < 0.001$) for PFS; and metastasis ($p = 0.01$) for M-PFS. The 'missing-data modality' approach ($N = 504$) led to different associations, including significance reached due to variables with the modality 'missing'. The CC analysis led to different results and reduced precision.

Conclusion: The CC population was not representative of the eligible population, introducing bias, in addition to worst precision. The 'missing-data modality method' results in biased estimates in non-randomized studies, as outcomes may be related to variables with missing values. Appropriate statistical methods for missing covariates, for example, MI, should therefore be considered.

Keywords: competing risks, irradiation, sarcoma, survival analysis

Received: 1 September 2023; revised manuscript accepted: 29 November 2023.

Correspondence to:
Carine Bellera
INSERM CIC1401, Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, 229 Cours de l'Argonne, Bordeaux 33076, France

Univ. Bordeaux, INSERM, Bordeaux Population Health Research Center, Epicene team, UMR 1219, Bordeaux, France
C.bellera@bordeaux.unicancer.fr

Noémie Huchet
Antoine Giraud
INSERM CIC1401, Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France

Nicolas Penel
Department of Medical Oncology, Centre Oscar Lambret, Lille, France

Lille University, Lille, France

Sylvie Bonvalot
Surgery Department, Institut Curie, Comprehensive Cancer Center, Paris, France

Juliette Thariat
Centre François Baclesse, Comprehensive Cancer Center, Caen, France

Laboratoire de physique Corpusculaire IN2P3/ENSICAEN/CNRS UMR 6534, Normandie Université, Caen France

Françoise Ducimetière
Jean-Yves Blay
Department of Medical Oncology, Centre Léon Bérard, Comprehensive Cancer Center, Lyon, France

Introduction

Observational studies can bring information on patient profiles not included in randomized clinical trials (RCT) and supplement with real-life knowledge on patient management, treatment strategies, and long-term survival. They complement the results of RCT by allowing one to assess the generalizability of survival outcomes reported

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the Sage and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Maud Toulmonde
Department of Medical Oncology, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France

Axel Le Cesne
Department of Medical Oncology, Gustave Roussy Cancer Campus, Comprehensive Cancer Center, Villejuif, France

in RCT to the real-life setting, to expand the generalizability of trials' results to underrepresented populations (e.g. rare diseases), and to generate scientific hypotheses.

Although radiation therapy is one of the available treatments for patients with cancer, it is also a risk factor for secondary tumors including soft-tissue sarcomas (STSs).[1] STSs are rare tumors that represent a heterogeneous group of diseases accounting for 1% of all malignancies in adults.[2] Sarcoma in irradiated fields (SIF) represent about 1–2% of all STSs; their multifactorial physiopathology is largely not understood.[1,3,4] Given the low incidence of SIF, limited data are available regarding treatment outcomes in this population. Prospective clinical trials are hardly feasible in the case of SIF or require international effort and a very long recruiting period. Large multicenter observational studies can thus provide valuable and irreplaceable information in this specific setting.

The French National Cancer Institute (INCa) funded a clinical network for sarcoma (NETSARC network) in 2009, to improve the management and outcome of sarcoma patients.[5] In all, 26 reference centers throughout the nation were identified. A network for expert pathology diagnosis in sarcoma (RRePS) gathering 23 reference centers for pathology in charge of the second bio-pathological opinion for each suspected case was also created. A common database (netsarc.org) gathering all cases of sarcoma presented to the multidisciplinary tumor board (MDTB) was created and implemented, collecting data on the diagnostic, therapeutic management, and clinical outcomes in terms of relapse and survival. This database includes both cases managed within the NETSARC network and those managed outside this network, and in the latter cases, the collected data are much less precise. Nevertheless, this database led to several publications improving significantly the scientific knowledge of sarcomas,[5,6] including rare histologic subtypes.[7] This database thus represents a unique opportunity to provide a better understanding of clinical outcomes in patients with SIF.

Missing data is a pervasive problem in both experimental and observational medical research, causing a loss of information and potentially biasing inferences.[8] Missing data in covariates is a problem in many survival studies and can render estimators biased when analyses are restricted to the population with complete information only as the restricted population may not be representative of the target population, or can lead to a loss of power to detect associations between explanatory variables and time-to-event endpoints. In these conditions, appropriate statistical methods that properly account for missing covariates should be applied.

The aim of the present study was to assess prognostic factors of overall survival (OS), progression-free survival (PFS), and metastatic progression-free survival (MPFS) in patients with SIF, based on the observational retrospective NETSARC database and by properly accounting for missing covariates.

Patients and methods

The NETSARC database

Collected parameters, as well as the strict quality insurance procedures, can be found in previous publications on the NETSARC database.[5,7] The NETSARC database allows (i) to exhaustively describe the incident and prevalent population of sarcoma patients in France, by cross-comparison of the pathological review database (rreps.org) and of the clinical database (netsarc.org), (ii) to monitor the diagnostic and initial treatment procedures, and (iii) to monitor patient outcome in particular survival and relapse. The database includes a limited set of data, describing patients and tumor characteristics, surgery, relapse, and survival. The following data were systematically collected: (1) tumor characteristics (histological subtypes, primary location, depth, lymph node involvement, or metastasis at diagnosis), (2) patient characteristics (sex, age, prior history of cancer, prior history of radiation therapy, preexisting lymphedema, known genetic predisposing conditions, human immunodeficiency virus infection, and grade according to the French Federation of Cancers Sarcoma Group), (3) management characteristics (initial management at the reference center, surgery performed, surgery quality, [neo]adjuvant radiotherapy or chemotherapy, and complete remission at the end of initial management), and (4) outcome (occurrence of local or distant relapse and status at last follow-up). The term 'non-gradable' means that the prognostic value of the FNCLCC grading system is not established for the considered histopathological type, even if technically one can describe the mitotic count, the necrosis, and the differentiation.

Eligibility criteria

The eligibility criteria of the present study were patients with soft-tissue or visceral SIF surgery of the primary tumor and a history of previous cancer. SIF was defined as follows: history of radiation exposure at least 3 years before the development of sarcoma,[9] occurrences of STS within the radiation field, and pathologic confirmation of a sarcoma that is histologically different from primary cancer.

Survival endpoints

OS was defined as the time interval between the date of initial sarcoma diagnosis and death (any cause). PFS was defined as the time interval between the date of initial sarcoma diagnosis and progression (local or distant) or death, whichever came first. MPFS was defined as the time interval between the date of initial sarcoma diagnosis and distant progression or death (any cause), whichever came first, as per DATECAN.[10]

Prognostic factors

Age (<25, 25–49, 50–74, and 75+) and sex were considered as potential prognostic factors. Clinical characteristics of the tumor included tumor site (soft tissue, viscera), tumor size (<5 cm, 5–10 cm, and >10 cm), depth of the tumor (superficial, deep), grade of the tumor (1, 2, 3, non-gradable), as well as the presence of metastases at diagnosis (yes, no). Pre-surgical imaging (yes, no), pre-surgical biopsy (yes, no), surgical resection margins (R0, R1, and R2), expertise of the surgeon (surgeon from NETSARC network, surgeon specialized in STSs outside network, and surgeon from outside network), and neo-adjuvant treatment (yes, no) were also investigated.

Statistical analysis

The eligible population involved all patients of the NETSARC database satisfying eligibility criteria with information available regarding survival outcomes (events and dates available). Complete cases were defined as eligible patients with information available for all prognostic factors.

Qualitative variables were described using counts and proportions. Median follow-up time was estimated using reverse Kaplan–Meier.[11] OS and PFS were described using the Kaplan–Meier estimator; median survival times were reported with a 95% confidence interval (95% CI). MPFS was described using the Aalen-Johansen estimator to account for competing risk (local progression); median cumulative incidence was reported with 95% CI.

OS and PFS were modeled using Cox proportional hazards models, and hazard ratios (HR) were reported to measure association with candidate prognostic factors, together with their 95% CI. MPFS was modeled using a Fine and Gray model to account for the presence of local progression, considered as a competing event. The model allows one to estimate the sub-distribution hazard function, for a given type of event (here distant progression or death), defined as the instantaneous rate of occurrence of the given type of event in subjects who have not yet experienced an event of that type. The Fine-Gray sub-distribution hazard model estimates the effect of covariates on the sub-distribution hazard function.[12]

Multivariate modeling strategy for survival outcomes was based on the following steps: (i) assessment of the correlation between candidate prognostic factors, (ii) univariate modeling, (iii) selection of prognostic factors to be included in the full multivariate model ($p < 20\%$), (iv) model reduction based on a manual backward selection process to account for potential confounder and effect modifier, and (v) investigation of potential interactions. We assessed model adequacy and ensured that the hypothesis of proportional hazards (PH) was not violated. In case of PH violation, we partitioned the time axis and reported distinct HR for each time period.

We accounted for the presence of missing prognostic factors by relying on a multiple imputation (MI) approach. If the missingness of a variable is related to observed characteristics but not to unobserved characteristics, the data are assumed 'missing at random' (MAR).[13] In such a case, the observed data can be used to estimate the missing value and subsequently replace (impute) the missing value by that estimate. This is done using a multivariable regression model, which imputes the missing value with the most likely value, based on all observed patient characteristics, including the outcome. MI involves 'filling in' each missing value withdraws from an appropriate distribution, leading to a number N_D of completed datasets. The substantive model (e.g. the Cox PH model for the analysis of OS) is then fitted to each of the N_D completed datasets, and the results are combined

across the N_D datasets, while accounting for the uncertainty because the imputed values were not actually observed, but rather estimated. We relied on imputation by fully conditional specification (FCS).[8] FCS MI involves specifying a series of univariate models for the conditional distribution of each partially observed variable given the other variables. FCS-MI was fitted using the R package smcfcs. With the *MI approach*, all patients are included in the analyses.

An alternative approach for handling missing covariates, easy to implement but potentially biased, is the *complete-case analysis*, which has been reported to be used in more than half of observational time-to-event studies in oncology.[14] We applied this second approach and thus omitted from the analysis patients with any missing prognostic factor. With the *complete-case analysis*, only patients with all prognostic factors available are analyzed.

Finally, another popular and simple approach for dealing with missing covariates is to replace the missing observations in a covariate with the mean or median value for a quantitative covariate, or the use of a missing indicator category for categorical covariates. We thus applied this third method and created a dedicated category for missing values for prognostic factors (all qualitative data in our situation), for example, tumor size was considered as a 4-modality variable in the statistical analyses: <5 cm, 5–10 cm, >10 cm, and missing. We will refer to this approach as the *missing-category approach* thereafter. With the *missing-category approach*, all patients are included in the analyses.

Subgroup analyses were conducted in patients with angiosarcoma, the most frequent histological type in our population.

Results

Between 1 January 2010 and 31 December 2017, a total of 17,684 adult patients with soft tissue or visceral sarcoma and surgery of the primary tumor were included in the database. Of those, 504 patients with SIF were eligible, including 167 complete cases (CC; 33%). In the eligible set, more than 20% presented with missing data for surgical resection margins, pre-surgery imaging, or neo-adjuvant treatment, and more than 10% of the patients presented with missing data for tumor size or tumor depth (Supplemental Table 1). Angiosarcomas represented the vast majority of SIF (42%).

Patient, tumor, and treatment characteristics are summarized in Table 1. As compared to the whole eligible population, CC presented more often with STSs (78% *versus* 69%) and grade I–II tumors (32% *versus* 21%).

The median OS was 7.4 years and 6.2 years for the eligible and CC populations, respectively (Table 2). Final multivariate models for the analysis of prognostic factors of OS are provided in Table 3 (univariate models available in Supplemental Table 2). Using MI, the presence of metastases at diagnosis was associated with short OS [HR = 1.83; 95%CI: (1.06; 3.45), $p = 0.04$]. A similar significant association was found for R1/R2 surgical margins as compared to R0 margins ($p < 0.001$), with an increasing risk over time (identified following investigation of Schoenfeld residuals): before 4 years reported estimates were $HR_{R1/R0} = 1.07$ [95%CI: (0.61; 1.8886)] and $HR_{R2/R0} = 2.40$ [95%CI: (1.07; 5.34)] while estimates after 4 years were $HR_{R1/R0} = 3.58$ (95%CI: [1.69; 7.55]) and $HR_{R2/R0} = 4.42$ [95%CI: (1.44; 13.64)]. The analysis based on the *missing-category approach* led to similar associations for surgical margins but no association was found for metastases at diagnosis. The *complete-case analysis* revealed that visceral (as compared to soft tissue) tumors, R1/R2 resection (as compared to R0), and surgery performed outside the referral center were associated with shorter OS.

The median PFS was 1.5 years and 2.0 years for the eligible and CC populations, respectively (Table 2). Final multivariate models for the analysis of prognostic factors of OS are provided in Table 4 (univariate models available in Supplemental Table 3). Using MI, R1, and R2 surgical margins as compared to R0 margins ($p < 0.001$), as well as grade II and III and non-gradable tumors ($p = 0.002$) were associated with shorter PFS. The association with surgical margins was also found using the missing-category analysis, which also revealed associations between the size of the tumor and the expertise of the surgeon. Of note, the 95%CI for the hazard ratio for the tumors with a missing size was the only one that did not include the null value [HR = 1.91; 95% CI: (1.29; 2.82)], while 95%CI for HR for tumors of size 5–10 cm to greater than 10 cm did both include the null value. Similarly, the 95% CI

Table 1. Characteristics of the patients with sarcoma in irradiated fields in the eligible population ($N = 504$) and in the complete-case population ($N = 167$).

Characteristics	Eligible population *n* (%)	Complete-case population *n* (%)
Age (years)		
<25	6 (1.2)	4 (2.4)
50–49	45 (8.9)	11 (6.6)
50–74	296 (58.7)	96 (57.3)
>74	157 (31.2)	56 (33.5)
Missing	–	–
Sex		
Female	405 (80.4)	133 (79.6)
Male	99 (19.6)	34 (20.4)
Missing	–	–
Site of the tumor		
Soft tissue	348 (69.0)	131 (78.4)
Viscera	156 (31.0)	36 (21.6)
Missing	–	–
Size of the tumor (cm)		
<5	207 (41.1)	73 (43.7)
5–10	187 (37.1)	73 (43.7)
>10	57 (11.3)	21 (12.6)
Missing	53 (10.5)	–
Depth of the tumor		
Deep	286 (56.7)	106 (63.5)
Superficial	163 (32.3)	61 (36.5)
Missing	55 (10.9)	–
Grade of the tumor		
Grade I	15 (3.0)	12 (7.2)
Grade II	90 (17.9)	41 (24.6)
Grade III	113 (22.4)	41 (24.6)
Not gradable	256 (50.8)	73 (43.7)
Missing	30 (6.0)	–
Metastases at diagnosis		
Yes	25 (5.0)	5 (3.0)
No	462 (91.7)	162 (97.0)
Missing	17 (3.4)	–
Pre-surgical imaging		
Yes	378 (75.0)	156 (93.4)
No	23 (4.6)	11 (6.6)
Missing	103 (20.4)	–
Pre-surgical biopsy		
Yes	387 (76.8)	144 (86.2)
No	89 (17.7)	23 (13.8)
Missing	28 (5.6)	–
Surgical resection margins		
R0	261 (51.8)	109 (65.3)
R1	115 (22.8)	51 (30.5)
R2	26 (5.2)	7 (4.2)
Missing	102 (20.2)	–
Center for surgical management		
Surgeon from the NETSARC network	235 (46.6)	85 (50.9)
Specialized surgeon outside the network	43 (8.5)	18 (10.8)
Surgeon outside network	214 (42.5)	64 (38.3)
Missing	12 (2.4)	–
Neo-adjuvant treatment		
Yes	74 (14.7)	33 (19.8)
No	247 (49.0)	134 (80.2)
Missing	183 (36.3)	–

for the hazard ratio for tumors with surgery performed by a surgeon with unknown/missing expertise did not include the null value. For the

Table 2. Survival outcomes of the patients with sarcoma in irradiated fields in the eligible population (*N* = 504) and in the complete-case population (*N* = 167).

Survival outcomes	Eligible population	Complete-case population
OS		
Median OS, 95%CI	7.42 years (6.59; 7.84)	6.15 years (5.77; not reached)
1-year OS rate	93.2% (90.9%; 95.7%)	90.6% (85.8%; 95.6%)
5-year OS rate	68.9% (63.5%; 74.8%)	71.2% (62.5%; 81.1%)
PFS		
Median PFS, 95%CI	1.51 years (1.34; 1.86)	2.02 years (1.37; 2.51)
1-year PFS rate	64.8% (60.3%; 69.6%)	72.0% (65.0%; 79.7%)
5-year PFS rate	21.5% (17.3%; 26.9%)	25.9% (18.8%; 35.7%)
MPFS		
1-year cumulative incidence	15.7% (12.2%; 19.2%)	11.6% (6.4%; 16.8%)
5-year cumulative incidence	33.3% (28.3%; 38.4%)	25.9% (18.0%; 33.7%)

95% CI, 95% confidence interval; MPFS, metastases progression-free survival; OS, overall survival; PFS, progression-free survival.

complete-case analysis, only the presence of metastases and the expertise of the surgeon were associated with shorter PFS.

The 5-year cumulative incidence for MPFS was 33% and 26% for the eligible and CC populations, respectively (Table 2). Final multivariate models for the analysis of prognostic factors of MPFS are provided in Table 5 (univariate models available in Supplemental Table 4). MI revealed increased risk in case of metastases at diagnoses [HR = 2.35; 95% CI: (1.22; 4.53)]. No association was found with the *missing-modality* approach. In the subgroup of complete cases, males, visceral tumors, metastases at diagnosis, and absence of pre-surgery biopsies were associated with poorer outcomes.

Angiosarcoma patients accounted for 42% of all eligible patients. Descriptive statistics as well as multivariate analyses for survival outcomes are reported (Supplemental Tables 5–10). Although results should be interpreted with caution due to the reduced sample size, the prognostic role of R1/R2 surgical margins as compared to R0 margins is worth mentioning. It was significantly associated with all survival outcomes in univariate models but this effect could be observed for multivariate analyses only for M-PFS.

Discussion

The aim of the present study was to assess prognostic factors of OS, PFS, and MPFS in patients with SIF, based on the observational retrospective NETSARC database and by properly accounting for missing covariates.

MI is an increasingly popular method for handling missing data which involves replicating the original dataset multiple times and, in each replication, replacing the missing values with plausible observations drawn from the posterior predictive distribution. MI is most often applied under the MAR assumption, which stipulates that the probability that data are missing is independent of the missing values, conditional on the observed data, although MI can also be used when data are missing not at random.[8] In the context of survival data, it remains difficult to recommend a specific imputation method as it will depend on the context of the study.[14] However, Bartlett's approach is recommended as the reference method.[8]

The missing-category approach might be appealing as it allows one to maintain statistical power. The resulting estimated association between the prognostic factor under study and outcome (e.g. OS) is a weighted average of two associations representing on one hand, the association between

Table 3. Prognostic factors of overall survival for patients with sarcoma in irradiated fields: final multivariate models.

Prognostic factors	Multiple imputation analysis ($n = 504$)		Missing-category analysis ($n = 504$)		Complete-case analysis ($n = 167$)	
	HR (95%CI)	*p* Value	HR (95%CI)	*p* Value	HR (95%CI)	*p* Value
Site of the tumor (Ref: Soft tissue)						
Viscera	–	NS	–	NS	0.37 [0.15; 0.91]	0.017
Metastases at diagnosis (Ref: No)						
Yes	1.83 [1.06; 3.45]	0.042	–	NS	–	NS
Missing	N/A					
Surgical resection margins (Ref: R0)						
R1	(*)	(*)	(*)	(*)	1.04 [0.53; 2.03]	0.026
R2					6.40 [2.05; 20.02]	
Missing					N/A	
Surgical resection margins (Ref: R0)						
Before 4 years		<0.001		0.002	(**)	(**)
R1	1.07 [0.61; 1.86]		0.97 [0.57; 1.65]			
R2	2.40 [1.07; 5.34]		2.30 [1.22; 4.67]			
Missing	N/A		0.90 [0.51; 1.59]			
After 4 years						
R1	3.58 [1.69; 7.55]		4.18 [1.95; 8.97]			
R2	4.42 [1.44; 13.64]		5.01 [1.61; 15.91]			
Missing	N/A		1.94 [0.84; 4.50]			
Center for surgical management (Ref: Surgeon from the NETSARC network)						
Specialized surgeon outside the network	–	NS	–	NS	1.22 [0.40; 3.67]	0.009
Surgeon outside network					2.76 [1.43; 3.54]	
Missing					N/A	

(*): Given the presence of a time-varying effect for this variable (i.e. non constant HR over time), HRs are reported for specific time windows (see subsequent line).
(**): Given the absence of a time-varying effect for this variable, HRs are reported globally (see previous line) and not for specific time windows.
95% CI: 95% confidence interval; HR: hazard ratio; N/A: not applicable; NS: not statistically significant.

the covariate and outcome, adjusted for all covariates, among the participants for whom all data were observed; and, on the other hand, the association between the covariate and outcome, adjusted only for complete covariates, among the participants for whom the covariate was not observed.[13] For nonrandomized studies, the second association will typically be biased because it is only partially adjusted for confounding. In addition, the first association is based on a complete-case analysis, so this association is unbiased only if missingness is conditionally independent

Table 4. Prognostic factors of progression-free survival for patients with sarcoma in irradiated fields: final multivariate models.

Prognostic factors	Multiple imputation analysis ($n = 504$)		Missing-category analysis ($n = 504$)		Complete-case analysis ($n = 167$)	
	HR (95%CI)	*p* Value	HR (95%CI)	*p* Value	HR (95%CI)	*p* Value
Size of the tumor (Ref: <5 cm)						
5–10 cm	–	NS	1.21 [0.92; 1.58]	0.014	–	NS
>10 cm			1.38 [0.95; 2.01]			
Missing			1.91 [1.29; 2.82]			
Grade of the tumor (Ref: Grade I)						
Grade II	2.00 [0.87; 4.59]	0.002	2.24 [0.95; 5.28]	0.005	–	NS
Grade III	2.79 [1.23; 6.36]		3.21 [1.37; 7.52]			
Not gradable	2.97 [1.33; 6.63]		3.23 [1.41; 7.39]			
Missing	N/A		2.66 [1.03; 6.88]			
Metastases at diagnosis (Ref: No)						
Yes	–	NS	–	NS	3.04 [1.21; 7.62]	0.041
Missing					N/A	
Surgical resection margins (Ref: R0)						
R1	1.45 [1.08; 1.95]	<0.001	1.44 [1.07; 1.94]	0.002	–	NS
R2	3.09 [1.89; 5.05]		2.60 [1.54; 4.42]			
Missing	N/A		1.00 [0.72; 1.38]			
Center for surgical management (Ref: Surgeon from the NETSARC network)						
Specialized surgeon outside the network	–	NS	1.44 [0.95; 2.17]	0.026	1.79 [0.94; 3.40]	0.004
Surgeon outside the network			1.37 [1.06; 1.77]		2.04 [1.32; 3.14]	
Missing			2.99 [1.07; 8.37]		N/A	

95% CI, 95% confidence interval; HR, hazard ratio; N/A, not applicable; NS, not statistically significant.

of the outcome. Given the nature of nonrandomized studies, in which covariates are commonly mutually related, this approach will almost always give biased results.[15] The missing-category method can thus be biased, inefficient, or underestimate the variance of estimates.[14]

Finally, although the *complete-case analysis* results in a loss of statistical power, it generally gives unbiased estimates when the participants without complete observations are a representative subset of the study population a situation known as 'missing completely at random (MCAR)'.[13] This situation however is rarely encountered and can be difficult to prove, and the direction of the bias (i.e. under- or over-estimation of the point estimates) is difficult to assess. In the present example, the population of complete cases was clearly not representative of the full sample, as CC presented more often with soft-tissue sarcoma and grade I–II tumors. Although this easy-to-implement method for handling missing data has been reported to be

Table 5. Prognostic factors of metastases progression-free survival for patients with sarcoma in irradiated fields: final multivariate models.

Prognostic factors	Multiple imputation analysis ($n = 504$)		Missing-category analysis ($n = 504$)		Complete-case analysis ($n = 167$)	
	sHR (95% CI)	*p* Value	sHR (95% CI)	*p* Value	sHR (95% CI)	*p* Value
Sex (Ref: Male)						
Female	–	NS	–	NS	0.41 [0.20; 0.84]	0.014
Site of the tumor (Ref: Soft tissue)						
Viscera	–	NS	–	NS	3.79 [1.79; 8.01]	<0.001
Metastases at diagnosis (Ref: No)						
Yes	2.35 (1.22; 4.53)	0.011	–	NS	11.33 [3.89; 33.02]	<0.001
Missing	N/A				N/A	
Pre-surgical biopsy (Ref: No)						
Yes	–	NS	–	NS	0.43 [0.19; 0.96]	0.038
Missing					N/A	

95% CI, 95% confidence interval; HR, hazard ratio; N/A, not applicable; NS, not statistically significant.

used in more than half of observational time-to-event studies in oncology, its use should therefore be discouraged, unless one can provide strong arguments in favor of the MCAR setting.[14]

This series is likely representative of SIF which represents 1.3% of STSs recoded in the database. In the end, the prognostic analysis demonstrates that prognostic factors seen in SIF are like prognostic factors for STSs. We stressed the importance of two intrinsic prognostic factors, grade, and presence of metastasis. We did not find that angiosarcoma in irradiated fields had a different outcome compared to other SIFs. This study underlined the importance of two extrinsic prognostic factors, quality of surgical margins (R0 resection), and center for surgical management. In this series, we did not observe the clinical benefit of neoadjuvant treatment.[5] In the specific cases of SIF sarcoma, preoperative radiation therapy is rarely done because of causing role of previous radiation therapy and because of the usual fibrotic aspect of surrounding tissue. As a consequence, preoperative chemotherapy could be discussed; nevertheless, a large part of patients had been exposed to anthracycline for the management of prior cancer (e.g. breast cancer or lymphoma). The role of neoadjuvant treatment remains a matter of debate in localized STSs[16] In major clinical trials assessing the role of neoadjuvant chemotherapy, patients with SIF had been excluded because of prior history of cancer. So, the role of preoperative chemotherapy in SIF remains an open question.[17]

The study limitations are inherent to its retrospective nature, with the critical issue of missing or imprecise data. As an example, the nature of the neoadjuvant was not collected; nevertheless, in the context of SIF, this neoadjuvant treatment is mostly preoperative chemotherapy rather than preoperative radiotherapy. We have considered that angiosarcoma is non-gradable since whatever the mitotic count, whatever the necrosis, whatever the differentiation, angiosarcoma must be regarded as an aggressive tumor.[18] The FNCLCC is less informative for this particular histological subtype. Table 4 clearly shows that the risk of relapse was similar in grade III SIF and non-gradable SIF which are mainly represented by angiosarcomas. To the best of our knowledge, the present study is one of the largest studies on prognostic factors of SIF. Nevertheless, subgroup analysis (e.g. prognosis angiosarcoma in irradiated fields) must be interpreted with caution (Supplemental Data).

Conclusion

In cases where retrospective studies constitute one of the best levels of evidence available (e.g. rare pathologies or exceptional patient populations), appropriate methods should be used to take missing data into account, to limit biases as much as possible. Working only on complete cases, better documented and better described by referral centers creates a selection bias, as illustrated in the present study. Consequently, the results of prognostic models vary greatly from one population to another, from one method of imputation to another. This is of major importance since missing data is inherent to retrospective studies, and more and more 'real-life-studies' are published. Physicians should pay attention to these issues when interpreting data.

Author's note

The present work was presented as a poster communication at the 2022 ASCO meeting.

Declarations

Ethics approval and consent to participate

In this retrospective study, all methods were carried out in accordance with relevant guidelines and regulations. Specifically, the data collection and further analysis were approved by the ethics committees as required by the applicable national legislation: approval by the Advisory Committee on Health Research Information Processing ('Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé', CCTIRS) on 16th September 2010, authorization number 10.403 and approval by the Information & Rights protection Committee ('Comité National Informatique et Liberté', CNIL) on 15th July 2013, authorization number 910390, Decision DR-2013-383. Signed informed consent was obtained from all study participants before registration.

Consent for publication

Not applicable.

Author contributions

Noémie Huchet: Formal analysis; Methodology; Software; Writing – review & editing.

Nicolas Penel: Conceptualization; Investigation; Methodology; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Sylvie Bonvalot: Investigation; Validation; Writing – review & editing.

Juliette Thariat: Investigation; Validation; Writing – review & editing.

Françoise Ducimetière: Investigation; Validation; Writing – review & editing.

Antoine Giraud: Investigation; Validation; Writing – review & editing.

Maud Toulmonde: Investigation; Validation; Writing – review & editing.

Axel Le Cesne: Investigation; Validation; Writing – review & editing.

Jean-Yves Blay: Investigation; Validation; Writing – review & editing.

Carine Bellera: Conceptualization; Formal analysis; Methodology; Project administration; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Acknowledgements

The authors would like to thank all the sarcoma teams and leaders of the French National Cancer Institute (INCa) for the continuous support to the project.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by grants from the French Cancer Institute (INCa) and Direction Générale de l'Offre de Soins (DGOS) (grant INCa_6291; NetSarc, InterSarc, and RrePS grants), the Labeled Clinical Research Consortium (Bordeaux Integrative Oncology Research [BRIO] and Lyon Integrative Cancer Research Program [LYric]; grant DGOS-INCa 4664), LabEx DEVweCAN (grant ANR-10-LABX-0061), European Clinical Trials in Rare Sarcomas (EUROSARC; grant FP7-278742), and Ligue de L'Ain contre le Cancer and la Fondation ARC.

The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Competing interests

The authors declare that there is no conflict of interest.

Availability of data and materials

The data that support the findings of this study are available from the NETSARC network but restrictions apply to the availability of these data. Data are available from upon reasonable request and with permission of the NETSARC network (https://netsarc.sarcomabcb.org/).

ORCID iDs

Nicolas Penel https://orcid.org/0000-0001-5243-1548

Carine Bellera https://orcid.org/0000-0001-7926-4671

Supplemental material

Supplemental material for this article is available online.

References

1. Mirjolet C, Diallo I, Bertaut A, *et al.* Treatment related factors associated with the risk of breast radio-induced-sarcoma. *Radiother Oncol* 2022; 171: 14–21.
2. Coindre JM, Terrier P, Guillou L, *et al.* Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group. *Cancer* 2001; 91: 1914–1926.
3. Gregersen PA, Olsen MH, Urbak SF, *et al.* Incidence and mortality of second primary cancers in Danish patients with retinoblastoma, 1943-2013. *JAMA Netw Open* 2020; 3: e2022126.
4. Goy E, Tomezak M, Facchin C, *et al.* The out-of-field dose in radiation therapy induces delayed tumorigenesis by senescence evasion. *Elife* 2022; 11: e67190.
5. Blay JY, Honoré C, Stoeckle E, *et al.*; NETSARC/REPPS/RESOS and French Sarcoma Group–Groupe d'Etude des Tumeurs Osseuses (GSF-GETO) Networks. Surgery in reference centers improves survival of sarcoma patients: a nationwide study. *Ann Oncol* 2019; 30: 1143–1153.
6. Boughzala-Bennadji R, Stoeckle E, Le Péchoux C, *et al.* Localized myxofibrosarcomas: roles of surgical margins and adjuvant radiation therapy. *Int J Radiat Oncol Biol Phys* 2018; 102: 399–406.
7. Penel N, Coindre JM, Giraud A, *et al.* Presentation and outcome of frequent and rare sarcoma histologic subtypes: a study of 10,262 patients with localized visceral/soft tissue sarcoma managed in reference centers. *Cancer* 2018; 124: 1179–1187.
8. Bartlett JW, Seaman SR, White IR, *et al.* Alzheimer's Disease Neuroimaging Initiative. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Methods Med Res* 2015; 24: 462–487.
9. Berrington de Gonzalez A, Gilbert E, Curtis R, *et al.* Second solid cancers after radiation therapy: a systematic review of the epidemiologic studies of the radiation dose-response relationship. *Int J Radiat Oncol Biol Phys* 2013; 86: 224–233.
10. Bellera CA, Penel N, Ouali M, *et al.* Guidelines for time-to-event end point definitions in sarcomas and gastrointestinal stromal tumors (GIST) trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol* 2015; 26: 865–872.
11. Schemper M and Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996; 17: 343–346.
12. Austin PC and Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med* 2017; 36: 4391–4400.
13. Groenwold RHH, White IR, Donders ART, *et al.* Missing covariate data in clinical research: when and when not to use the missing-indicator method for analysis. *CMAJ* 2012; 184: 1265–1269.
14. Carroll OU, Morris TP and Keogh RH. How are missing data in covariates handled in observational time-to-event studies in oncology? A systematic review. *BMC Med Res Methodol* 2020; 20: 134.
15. Donders AR, van der Heijden GJ, Stijnen T, *et al.* Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; 59: 1087–1091.
16. Spałek MJ, Kozak K, Czarnecka AM, *et al.* Neoadjuvant treatment options in soft tissue sarcomas. *Cancers (Basel)* 2020; 12: 2061.
17. Gronchi A, Palmerini E, Quagliuolo V, *et al.* Neoadjuvant chemotherapy in high-risk soft tissue sarcomas: final results of a randomized trial from Italian (ISG), Spanish (GEIS), French (FSG), and Polish (PSG) Sarcoma Groups. *J Clin Oncol* 2020; 38: 2178–2186.
18. Guillou L, Coindre JM, Bonichon F, *et al.* Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma. *J Clin Oncol* 1997; 15: 350–362.

Visit Sage journals online journals.sagepub.com/home/tam

Sage journals