

Challenges of Using ICD-9-CM and ICD-10-CM Codes for Soft-Tissue Sarcoma in Databases for Health Services Research

by Lisa M. Hess, PhD; Yajun E. Zhu, MS; Tomoko Sugihara, MS; Yun Fang, MS; Nicholas Collins; and Steven Nicol, MD

Abstract

Objectives: Soft-tissue sarcoma (STS) is a heterogeneous group of rare solid tumors that arise from various soft tissues in the body, such as muscle, fat, nerves, and blood vessels. Current International Classification of Diseases (ICD) coding systems include a set of nonspecific codes for malignancies of connective and soft tissue (ICD-9-CM code 171 and ICD-10-CM code C49). The goal of this study was to evaluate the use of these codes for health services research involving patients with a diagnosis of this rare malignancy.

Methods: Two databases were utilized to explore ICD coding for STS: claims data from Truven MarketScan and electronic medical records (EMRs) from Flatiron Health. Eligible patients from claims data were those with at least two ICD-9-CM codes of 171.x on two different days between July 1, 2004, and March 30, 2014. The treatment patterns of these cases were evaluated for consistency with known therapeutic approaches for STS. Eligible patients from the Flatiron EMR system were those who received olaratumab (a drug indicated only for use in patients diagnosed with STS) after its US Food and Drug Administration approval in October 2016 through the end of the data set (November 2017). ICD-10-CM codes were evaluated for this known STS cohort.

Results: In claims data, 4,159 patients were eligible for inclusion. Although national treatment guidelines include only a limited number of drugs used to treat STS, 98 unique anticancer drugs were identified as being used to treat patients in a claims data cohort. Only 7.7 percent of patients had claims for doxorubicin-based therapy and 3.8 percent had claims for ifosfamide-based therapy as initial treatment for STS, despite these being a standard of care. In the EMR data, 350 patients were eligible; only 170 patients (48.6 percent) had any evidence in the database of a connective or soft-tissue ICD-10-CM malignancy code within 60 days before or after initiation of olaratumab.

Conclusions: ICD coding for STS using the “Malignant neoplasm of connective and soft tissue” code is not reliable as a method to identify patients diagnosed with STS. Although codes reflecting the primary site of disease may have clinical relevance, lack of consistency in ICD coding for the diagnosis and treatment of this disease is a limiting factor in the ability to conduct real-world observational research of this rare disease. In the absence of consistent use of this code, an algorithm needs to be developed and validated to accurately identify patients with STS in these databases.

Key Words: sarcoma; ICD-9-CM; real-world data; retrospective study

Introduction

The International Classification of Diseases (ICD) system was implemented in 1948 to standardize the coding of disease for both clinical and research purposes. Although the International Classification of Diseases, Tenth Revision (ICD-10) version has been available and used worldwide since 1994,^{1, 2} the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) system was used for coding patient diagnosis in both inpatient and outpatient settings in the United States until the ICD-10-CM system was implemented in October 2015.

Because of the international incorporation of ICD coding into routine clinical practice, health services scientists rely on ICD codes to conduct research using real-world data.³ The foundation for the conduct of health services and outcomes research is access to claims and electronic medical record (EMR) databases.^{4, 5} Real-world research relies on these data sources to evaluate how medicines, treatment strategies, and procedures perform in uncontrolled settings with typical patients and to answer questions in situations where randomized trials are not feasible.⁶ Some of the basic requirements of such data sets for research include the ability to identify a study cohort, the availability of variables to answer the study question, and lack of gaps in individual patient data during the study period.⁷

The ICD classification system is a critical component for cohort identification using diagnosis codes, which enables such research to be conducted and replicated in various data sets.⁸ Standardized coding algorithms, such as the Charlson Comorbidity Index, further allow for consistent comorbidity evaluation across studies, providing a uniform approach to evaluate conditions.⁹ However, not all diseases have straightforward ICD diagnosis codes. Therefore, accurate cohort identification for research that uses ICD coding is critical, as more than one code (or a combination of codes) could potentially reflect the disease of interest when the codes are entered into databases for billing and clinical records, not for research purposes. When claims or other clinical databases are used for health services and outcomes research, the researcher must be aware of the intent of the data, as the use of codes for reimbursement may not mirror the use of codes for clinical care or the codes that would be used in a research setting. Therefore, even across these databases, coding may not be consistent for the same patient. Additionally, coding may be affected by incomplete information available to the coder, the coder's experience, and the legibility of written records from the clinical team.¹⁰

Limited work has been conducted validating the use of ICD-9 or ICD-10 codes for the accurate diagnosis of several cancers,¹¹ and evidence suggests the limitations of ICD coding for cancer research, generally. A study comparing ICD-9 coding to cancer registry data found that ICD-9 coding overestimated cancer cases, with a 39 percent false-positive rate, representing cancer cases identified by a ICD-9 code that failed to correspond to a cancer diagnosis in the patient medical record.¹² This finding may be due to the use of the code in an evaluation to rule out cancer. However, for some cancer types, such as bone and joint or brain and nervous system malignancies, the number of diagnoses was higher using ICD-9 codes than in the medical record.¹³ A second study used an ICD-9 algorithm combining diagnosis codes with procedure codes for receipt of chemotherapy to better identify breast, colorectal, and lung cancer cases.¹⁴ The combination of these codes resulted in false-positive rates of 10.2 percent, 10.1 percent, and 23.6 percent and false-negative rates of 25.7 percent, 27.8 percent, and 22.3 percent for breast, colorectal, and lung cancers, respectively. Other researchers are working to develop better algorithms to correctly identify conditions under study.¹⁵

The lack of a standardized approach to select ICD codes and the lack of validated algorithms to identify a population with a particular condition across studies limits the ability to evaluate the body of research. While these gaps are known for common diseases such as breast, lung, and colorectal cancers, these problems may be more pronounced for diseases such as soft-tissue sarcoma (STS), which is relatively rare and is diagnosed in approximately 13,040 people in the United States each year.¹⁶ STS is a heterogeneous set of tumors including more than 50 histological subtypes that can arise in any part of the body where mesenchymal tissues exist (e.g., cartilage, muscle, fat, blood vessels).¹⁷ The most common locations where STSs are diagnosed include the legs, arms, trunk, retroperitoneum, and head and neck.¹⁸ While codes specific to malignant neoplasms of other connective and soft tissue (ICD-9-CM 171 and

ICD-10-CM C49) are available, there are also numerous codes for the possible tumor locations where STSs may be found. For example, a patient with a uterine leiomyosarcoma (one of the more common STS subtypes) may have a diagnosis code of sarcoma (ICD-9-CM 171/ICD-10-CM C49) or may have a diagnosis code for uterine cancer (ICD-9-CM 179/ICD-10-CM C55 or 182/C54).

Both codes are technically accurate for clinical purposes, but they limit the ability to conduct research due to the inability to differentiate sarcomas from other nonsarcoma cancers that occur in the same location in the body. In the case of a uterine tumor, it is much more common for an epithelial carcinoma of the uterus to be associated with this ICD code, and the ICD-9-CM and ICD-10-CM systems are unable to differentiate between these diseases. While the World Health Organization has developed an oncology-specific histological coding system (ICD-O), it is not incorporated into the vast majority of real-world databases, as it is not used for billing or patient care purposes during routine clinical care.¹⁹ The National Comprehensive Cancer Network (NCCN) Drugs and Biologics Compendium includes several ICD-10-CM codes that are allowable for STS, including malignancies of other connective and soft tissue (C49), peripheral nerves and autonomic nervous system (C47), corpus uteri (C54), and retroperitoneum and peritoneum (C48).²⁰ These three latter codes have not typically been applied to health services research of STS because they are unable to differentiate between a sarcoma and an epithelial tumor at the same location. While researchers would prefer the presence of C49 codes to specifically identify sarcoma cases, the reality of needs for reimbursement to achieve treatment goals may preclude this possibility.

This study was designed to evaluate the ability to define a STS cohort using ICD-9-CM and ICD-10-CM codes using two approaches: first, to evaluate the cohort identified using sarcoma-specific codes in claims data, and, second, to evaluate a known STS cohort through an EMR system to identify which ICD-10-CM codes are used by clinicians.

Methods

Data Sources

Two databases were utilized to explore ICD coding for STS: claims data from Truven MarketScan and EMR data from Flatiron Health. The Truven MarketScan databases contain patient-level inpatient, outpatient, and drug data from commercial, Medicaid, and employer-sponsored Medicare supplemental plans. The data are collected from approximately 350 different insurance companies and third-party administrators. The data files are organized by patient enrollment, medical, and pharmacy claims. The enrollment file contains information on age, gender, US geographical region, health insurance payer type, employment status of the policyholder, and monthly enrollment status. The medical claims include detailed records for hospital inpatient admissions and outpatient medical claims using ICD-9-CM and ICD-10-CM diagnosis and procedure codes, as well as Current Procedural Terminology Medical Code Set (CPT) and Healthcare Common Procedure Coding System (HCPCS) codes, dates and place of service, duration of hospital stays, and both plan and patient payment amounts. The pharmacy claims include the national drug code, therapeutic class, dispense dates, quantity and days supplied, and plan payment and patient copayment amounts.

The Flatiron Health EMR database includes a geographically diverse population of patients with cancer who visit an oncologist in the Flatiron network of community and academic cancer centers and clinics. The database contains more than 2 million active cancer patient records collected from more than 265 cancer centers across the United States.²¹ The data are refreshed monthly and include structured EMR data elements, such as patient demographics (gender, race, birth year, and state of residence), type of cancer facility visited (community versus academic), clinical diagnoses and procedures (using ICD-9 and ICD-10 coding), stage of cancer diagnosis, laboratory data, biomarker tests and results, medications ordered and/or administered, and dose in milligrams of each drug administered.

Eligibility Criteria

Eligible patients from Truven MarketScan claims data were those with at least two ICD-9-CM codes of 171.x on two different days at any time between July 1, 2004, and March 30, 2014. Two codes were required to minimize the rate of false-positive cases, and the time period selected allowed up to one year

of follow-up through the end of 2015, which was the end date of the available claims data at the time of analysis. The requirement for a baseline period and follow-up period ensured that sufficient data were available to evaluate the treatment period over time. Patients were required to have evidence of receiving anticancer therapy. To reduce the rate of potential false-positive STS cases, patients were excluded if they had ICD-9-CM codes of 170.x, 207.x, 204.x, 201.x, 203.x, 205.x, 176.x, 191.x, 200.x, 181.x, 202.x, 208.x, 206.x, or 162.x. This cohort served as the *sarcoma code cohort*.

Eligible patients from the Flatiron Health EMR data were those age 18 years or older who received olaratumab after its US Food and Drug Administration (FDA) approval on October 31, 2016, through the end of the data set (November 30, 2017, at the time of analysis). Olaratumab is approved only for use in patients diagnosed with STS.²³ This cohort served as the *known STS cohort*.

Analytic Plan

Lines of therapy refers to the treatments used to care for the patient throughout the course of disease, regardless of the stage of disease. In this study, the start of the first line of therapy was defined in the data by the first observed date of anticancer therapy. The regimen was defined as the set of drugs used within the first 21 days of treatment in a line of therapy. A patient was determined to have progressed to a subsequent line of therapy when new anticancer agents were added and drugs from the current line of therapy were discontinued. Discontinuation of agents would end a line of therapy, but a new line of therapy would not start until the initiation of new anticancer drugs. Augmentation (e.g., adding a biologic agent to a chemotherapy backbone) and partial discontinuation (e.g., stopping some but not all drugs in a treatment regimen) did not constitute a change in line of therapy.

Descriptive analyses were conducted using SAS version 9.2. Both the claims and EMR cohorts were described by clinical and demographic variables provided in each data set. Treatment patterns (e.g., anticancer drugs received as well as combinations and sequence of drugs received) of patients in the claims cohort were evaluated for consistency with known therapeutic approaches for STS. Data from the EMR known STS cohort were evaluated to identify what ICD-10-CM codes were assigned to patients who received olaratumab for treatment of STS.

Both the claims and EMR data sets used in this study are de-identified in accordance with 45 *Code of Federal Regulations* §46.102 and are HIPAA (Health Insurance Portability and Accountability Act) compliant.

Results

A total of 4,159 patients met eligibility criteria for the sarcoma code cohort, and 350 patients met eligibility criteria for the EMR known STS cohort data set (see Table 1). The mean age of patients was 63.2 years in the known STS cohort, compared with 58.9 years in the sarcoma code cohort. The geographic region was missing for approximately 23.6 percent of patients in the known STS cohort versus less than 8.7 percent in the sarcoma code cohort.

Sarcoma Code Cohort

The treatment patterns of the sarcoma code cohort are demonstrated in Table 2, with few regimens representative of therapies used for the treatment of STS. The most common drug combinations reported were nonspecific (unclassified) codes, which cannot be associated with any particular drug. Other common regimens are more likely to be used in nonmelanoma skin cancer (e.g., imiquimod), renal cell carcinoma (e.g., pazopanib), colorectal/gastric cancers (e.g., fluorouracil/capecitabine), breast cancers (e.g., anastrozole, tamoxifen, letrozole), or squamous head and neck tumors (e.g., cetuximab) than to treat STS, based on NCCN clinical treatment guidelines. This finding may suggest either that they are epithelial cancers miscoded as STS or that these are STS cases that are not treated in accordance with evidence-based guidelines. The claims database contains insufficient details associated with the data to determine which may be the case.

In Table 3, the NCCN-recommended therapies (doxorubicin- and ifosfamide-based treatment) are reported in only 7.7 and 3.8 percent of patients in claims data for the initial treatment for STS. In real-world settings, the use of these agents would be expected to be used in most patients. While comparison of these cohorts based on drug use is not appropriate given the selection criteria for the known STS cohort, the presence of agents and regimens that would not be expected to be used in the care of a patient with sarcoma is common in the claims data (e.g., single-agent fluorouracil, imiquimod, cetuximab, anastrozole, tamoxifen, letrozole, and pazopanib), whereas none of these drugs are found in the care of patients in the known STS cohort. Drugs that appear in the claims data may be used for some forms of sarcoma include imatinib (used for gastrointestinal stromal tumors) and interferon- α -2B (used for Kaposi's sarcoma). Olaratumab is not indicated for either of these forms of sarcoma.

Known STS Cohort

The ICD-10-CM neoplasm-related codes used in the known sarcoma cohort are presented in Table 4. Less than half of patients (48.6 percent) have a code for connective and soft tissue malignancies (C49.x) included in their record within 60 days before or after initiation of a known STS therapy. Other common codes included secondary malignancies of lung and bone (likely suggesting metastasized disease), and other primary cancer sites that are consistent with the site of origin of sarcoma (e.g., peritoneum, lung, uterus). Although the reported codes are consistent with what may occur in the course of STS, the frequency of any single code was low, with no identifiable primary site being reported in more than 5 percent of the known STS population (see Table 4). Consistent with the challenges of STS, many codes in Table 3 were not specific to any tumor location, suggesting that either the exact sarcoma diagnosis was unknown or no ICD-10-CM code could be identified to accurately categorize the patient's disease (e.g., many patients also had benign diagnoses and codes for undefined sites).

Discussion

This evaluation of claims data, based on ICD-9-CM coding, and of a known STS cohort, based on the use of a drug approved only for use in STS, has identified a number of challenges in the identification of sarcoma patients in real-world data sets. These challenges can be summarized as follows: low use of the ICD-9-CM/ICD-10-CM sarcoma code (171.x/C49.x), use of the ICD sarcoma codes for nonsarcomas (perhaps as a rule-out code), and lack of differentiation of either drugs or ICD-9-CM/ICD-10-CM codes for site-coded sarcomas versus other epithelial tumors that commonly occur in these sites. The Flatiron EMR data set including olaratumab users represented a known sarcoma cohort, which demonstrates the infrequency of use of the ICD-10-CM sarcoma code. The identification of patients in the claims database with the use of this code alone identifies only a subset of the target patient population and further is likely to include patients who are not true sarcoma patients. Although a known sarcoma cohort was identified using a specific drug name, this cohort is only a small subset of the larger sarcoma patient population and not representative of all sarcoma patients. The risks of misclassification in real-world databases include providing incorrect information about treatment patterns, treatment duration, and survival estimates.

In 2015 and 2016, the US FDA approved trebectedin and olaratumab, respectively.^{23, 24} Trebectedin is approved for use in patients with specific forms of sarcoma (liposarcoma or leiomyosarcoma, after failure of doxorubicin), whereas olaratumab is approved broadly for any STS subtype, which facilitated the identification of a broader STS cohort for ICD coding evaluation. Olaratumab, and to some extent trebectedin, provide new opportunities to identify sarcoma patients in real-world data sets and to begin to develop algorithms that can be validated in further research among patients diagnosed with this relatively rare and heterogeneous cancer. Unfortunately, at the time of analysis, neither trebectedin nor olaratumab had a J-code in the claims data set for analysis, and olaratumab had not yet been approved by the FDA for use in patients with STS during the time period of the claims data set. Future research should evaluate the use of claims data using these J-codes, when available, to develop and validate a claims cohort algorithm for sarcoma.

In prior work, researchers have attempted to overcome the limitations of ICD coding for STS by combining codes with drugs that are commonly used to treat sarcoma to reduce the bias.²⁵ While this approach is reasonable, the findings in this study suggest that this approach will not accurately identify a

complete cohort and will miss perhaps half of sarcoma cases that may not be associated with an expected ICD diagnosis code. This approach may minimize false positives, but it will also exclude true positives. Limiting any database study to sarcoma codes may exclude half of all true STS patients, many of whom may have a tumor-location ICD-10-CM code only.

Other researchers have attempted to combine nonsarcoma codes with drugs known to be used in sarcoma along with other site-specific ICD-9 codes.²⁶ Efforts to identify a cohort by limiting the search to include only drugs used to treat STS are challenged by the fact that many drugs used to treat sarcoma are also used to treat epithelial and other nonsarcoma cancers, and by the fact that the codes used for site-specific diseases may include epithelial and other nonsarcoma tumors. Prior studies of breast, colorectal, and lung cancers have shown that these approaches still have high false-positive and false-negative rates.²⁷ There is a critical need to define STS cohorts and to validate the approach to minimize the risk of misclassification or underrepresentation. Validation studies are needed to determine how well these algorithms perform. Until more standardized approaches are developed to ensure that specific ICD-10 codes are used when a patient is diagnosed with sarcoma, it is recommended that an algorithm be developed and tested by linking either claims or EMR data to patient medical records, where the diagnosis can be verified. Real-world data can provide a foundation for progress to be made in the care of patients diagnosed with STS, as recruitment for clinical trials involving rare diseases and tumor types can take many years.

Conclusion

The use of ICD-9 or ICD-10 codes to identify a sarcoma cohort for health services and outcomes research has limitations. Research using these algorithms should clearly state the limitations and barriers with regard to both misclassification and missed sarcoma cases using these approaches. In the absence of standardized ICD-10-CM coding for these diseases, researchers will need to develop and validate an ICD-10-CM coding algorithm that can be used across observational database studies of sarcoma to ensure that the findings are appropriately applied to clinical practice.

Lisa M. Hess, PhD, is a research advisor at Eli Lilly and Company and adjunct professor of medicine and public health at Indiana University in Indianapolis, IN.

Yajun E. Zhu, MS, is a statistician at Eli Lilly and Company in Indianapolis, IN.

Tomoko Sugihara, MS, is a principal statistical programmer of clinical solutions at Syneos Health in Indianapolis, IN.

Yun Fang, MS, is a principal statistical programmer of clinical solutions at Syneos Health in Indianapolis, IN.

Nicholas Collins is an associate consultant at Eli Lilly and Company in Indianapolis, IN.

Steven Nicol, MD, is a retired medical advisor at Eli Lilly and Company in Indianapolis, IN.

Notes

1. Rogers, C. “ICD-10: Was All the Delay Necessary?” *Health eSource* 13, no. 6 (2017).
2. World Health Organization. “Classifications.” 2016. Available at <http://www.who.int/classifications/icd/en/> (accessed February 14, 2018).
3. De Coster, C., H. Quan, A. Finlayson, et al. “Identifying Priorities in Methodological Research Using ICD-9-CM and ICD-10 Administrative Data: Report from an International Consortium.” *BMC Health Services Research* 6 (2006): 77.
4. Myers, L., and J. Stevens. “Using EHR to Conduct Outcome and Health Services Research.” In MIT Critical Data, *Secondary Analysis of Electronic Health Records*. Cambridge, MA: Springer, 2016, 61–70.
5. Schneeweiss, S., and J. Avorn. “A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics.” *Journal of Clinical Epidemiology* 58, no. 4 (2005): 323–337.
6. Motheral, B. R., and K. A. Fairman. “The Use of Claims Databases for Outcomes Research: Rationale, Challenges, and Strategies.” *Clinical Therapeutics* 19, no. 2 (1997): 346–66.
7. Schneeweiss, S., and J. Avorn. “A Review of Uses of Health Care Utilization Databases for Epidemiologic Research on Therapeutics.”
8. Goodman, R. A., S. F. Posner, E. S. Huang, A. K. Parekh, and H. K. Koh. “Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice.” *Preventing Chronic Disease* 10 (2013): E66.
9. Deyo, R. A., D. C. Cherkin, and M. A. Ciol. “Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Databases.” *Journal of Clinical Epidemiology* 45, no. 6 (1992): 613–19.
10. Hersh, W. R., M. G. Weiner, P. J. Embi, et al. “Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research.” *Medical Care* 51, no. 8, suppl. 3 (2013): S30–S37.
11. Whyte, J. L., N. M. Engel-Nitz, A. Teitelbaum, G. Gomez Rey, and J. D. Kallich. “An Evaluation of Algorithms for Identifying Metastatic Breast, Lung, or Colorectal Cancer in Administrative Claims Data.” *Medical Care* 53, no. 7 (2015): e49–e57.
12. Park, L. S., J. P. Tate, M. C. Rodriguez-Barradas, et al. “Cancer Incidence in HIV-Infected Versus Uninfected Veterans: Comparison of Cancer Registry and ICD-9 Code Diagnoses.” *Journal of AIDS and Clinical Research* 5, no. 7 (2014): 1000318.
13. Ibid.
14. Baldi, I., P. Vicari, D. Di Cuonzo, et al. “A High Positive Predictive Value Algorithm Using Hospital Administrative Data Identified Incident Cancer Cases.” *Journal of Clinical Epidemiology* 61, no. 4 (2008): 373–79.
15. Abraha, I., D. Serraino, G. Giovannini, et al. “Validity of ICD-9-CM Codes for Breast, Lung and Colorectal Cancers in Three Italian Administrative Healthcare Databases: A Diagnostic Accuracy Study Protocol.” *BMJ Open* 6, no. 3 (2016): e010547.
16. Siegel, R. L., K. D. Miller, and A. Jemal. “Cancer Statistics, 2018.” *CA: A Cancer Journal for Clinicians* 68, no. 1 (2018): 7–30.
17. Kransdorf, M. J. “Malignant Soft-Tissue Tumors in a Large Referral Population: Distribution of Diagnoses by Age, Sex, and Location.” *AJR American Journal of Roentgenology* 164, no. 1 (1995): 129–34.
18. Ibid.

19. Fritz, A., C. Percy, A. Jack, K. Shanmugaratnam, L. Sobin, D. M. Parkin, and S. Whelan. *International Classification of Diseases for Oncology (ICD-O)*. 3rd ed. Geneva, Switzerland: World Health Organization, 2013. Available at http://apps.who.int/iris/bitstream/10665/96612/1/9789241548496_eng.pdf (accessed February 15, 2018).
20. National Comprehensive Cancer Network (NCCN). *NCCN Drugs and Biologics Compendium: Soft Tissue Sarcoma 1*. 2018.
21. Flatiron. "About Us." 2018. Available at <https://flatiron.com/about-us/> (accessed February 15, 2018).
22. "Olaratumab Approved for Soft-Tissue Sarcoma." *Cancer Discovery*. 6, no. 12 (2016): 1297.
24. Ibid.
25. Barone, A., D. C. Chi, M. R. Theoret, et al. "FDA Approval Summary: Trabectedin for Unresectable or Metastatic Liposarcoma or Leiomyosarcoma Following an Anthracycline-Containing Regimen." *Clinical Cancer Research* 23, no. 24 (2017): 7448–53.
26. Villalobos, V. M., S. D. Byfield, S. R. Ghate, and O. Adejoro. "A Retrospective Cohort Study of Treatment Patterns among Patients with Metastatic Soft Tissue Sarcoma in the US." *Clinical Sarcoma Research* 7 (2017): 18.
27. Duh, M. S., M. D. Hackshaw, J. I. Ivanova, et al. "Costs Associated with Intravenous Cancer Therapy Administration in Patients with Metastatic Soft Tissue Sarcoma in a US Population." *Sarcoma* 2013 (2013): 947413.
28. Baldi, I., P. Vicari, D/ Di Cuonzo, et al. "A High Positive Predictive Value Algorithm Using Hospital Administrative Data Identified Incident Cancer Cases."

Table 1

Eligible Study Cohorts

Characteristic	Sarcoma Code Cohort (<i>n</i> = 4,159)	Known STS cohort (<i>n</i> = 350)
Gender, number (percent)		
Male	2,009 (48.3)	159 (45.4)
Female	2,150 (51.7)	191 (54.6)
Age in years, mean (SD)	58.9 (14.6)	63.2 (12.8)
Geographic region of United States, number (percent)		
Northeast	670 (16.1)	52 (14.9)
North Central/Midwest	908 (21.8)	48 (13.7)
South	1,442 (34.7)	128 (36.6)
West	777 (18.7)	35 (10.0)
Unknown/missing	362 (8.7)	87 (24.9)

Abbreviations: SD, standard deviation; STS, soft-tissue sarcoma.

Table 2

Treatment Regimens Used by More Than 1 Percent of the Sarcoma Code Cohort in First- or Second-Line Therapy, in Claims Data

Regimen	Number (Percent)	
	Line 1 (n = 4,159)	Line 2 (n = 1,339)
Unclassified drugs ^a	768 (18.47)	150 (11.20)
Fluorouracil	373 (8.97)	91 (6.80)
Docetaxel + gemcitabine	284 (6.83)	64 (4.78)
Cisplatin	266 (6.40)	28 (2.09)
Carboplatin + paclitaxel	149 (3.58)	35 (2.61)
Imiquimod	145 (3.49)	48 (3.58)
Imatinib	136 (3.27)	13 (0.97)
Doxorubicin	93 (2.24)	32 (2.39)
Doxorubicin + ifosfamide	83 (2.00)	20 (1.49)
Cetuximab	79 (1.90)	28 (2.09)
Fluorouracil + oxaliplatin	67 (1.61)	8 (0.60)
Gemcitabine	65 (1.56)	35 (2.61)
Anastrozole	60 (1.44)	33 (2.46)
Cisplatin + docetaxel + fluorouracil	58 (1.39)	3 (0.22)
Carboplatin + paclitaxel + unclassified drugs ^a	56 (1.35)	11 (0.82)
Interferon alfa-2B	55 (1.32)	2 (0.15)
Tamoxifen	55 (1.32)	18 (1.34)
Letrozole	54 (1.30)	20 (1.49)
Paclitaxel	49 (1.18)	31 (2.32)
Capecitabine	47 (1.13)	25 (1.87)
Carboplatin	22 (0.53)	37 (2.76)
Liposomal doxorubicin	32 (0.77)	33 (2.46)
Pazopanib	22 (0.53)	25 (1.87)
Docetaxel + gemcitabine + unclassified drugs ^a	22 (0.53)	15 (1.12)

^a Drugs with nonspecific HCPCS codes (e.g., J3490, J3590, J9999).

Table 3

Use of Doxorubicin and/or Ifosfamide

Drug Exposure	Number/Total (Percent)	
	Sarcoma Code Cohort	Known STS Cohort
Doxorubicin		
Line 1	320/4,159 (7.7)	216/350 (61.7)
Line 2	90/1,339 (6.7)	92/194 (47.4)
Line 3	32/481 (6.7)	36/79 (45.6)
Ifosfamide		
Line 1	159/4,159 (3.8)	14/350 (4.0)
Line 2	51/1,339 (3.8)	9/194 (4.6)
Line 3	13/481 (2.7)	1/79 (1.3)

Abbreviation: STS, soft-tissue sarcoma.

Table 4

Most Common ICD-10-CM Primary Neoplasm-related Codes (C00–D49) Used for the Known Soft-Tissue Sarcoma Cohort ($n = 350$)

Code	Number (Percent) ^a
C49, Malignant neoplasm of other connective and soft tissue	170 (48.6)
C78, Secondary malignant neoplasm of lung	69 (19.7)
C79, Secondary malignant neoplasm of bone	64 (18.3)
C48, Malignant neoplasm of retroperitoneum and peritoneum	23 (6.6)
C76, Malignant neoplasm of other and ill-defined sites	17 (4.9)
C34, Malignant neoplasm of bronchus and lung	10 (2.9)
C54, Malignant neoplasm of corpus uteri	10 (2.9)
C80, Malignant neoplasm without specification of site	9 (2.6)
C55, Malignant neoplasm of uterus, part unspecified	8 (2.3)
C22, Malignant neoplasm of liver and intrahepatic bile ducts	8 (2.3)
C50, Malignant neoplasm of breast	7 (2.0)
C41, Malignant neoplasm of bone and articular cartilage of other and unspecified sites	6 (1.7)
C40, Malignant neoplasm of bone and articular cartilage of limbs	5 (1.4)
C61, Malignant neoplasm of prostate	4 (1.1)
C44, Other and unspecified malignant neoplasm of skin	4 (1.1)
C47, Malignant neoplasm of peripheral nerves and autonomic nervous system	3 (0.9)
C53, Malignant neoplasm of cervix uteri	2 (0.6)
C71, Malignant neoplasm of brain	2 (0.6)
C46, Kaposi's sarcoma	2 (0.6)
C96, Other and unspecified malignant neoplasms of lymphoid, hematopoietic and related tissue	2 (0.6)
C64, Malignant neoplasm of kidney, except renal pelvis	2 (0.6)
C62, Malignant neoplasm of testis	1 (0.3)
C57, Malignant neoplasm of other and unspecified female genital organs	1 (0.3)
C25, Malignant neoplasm of pancreas	1 (0.3)
C26, Malignant neoplasm of other and ill-defined digestive organs	1 (0.3)
C00, Malignant neoplasm of lip	1 (0.3)
C69, Malignant neoplasm of eye and adnexa	1 (0.3)
C38, Malignant neoplasm of heart, mediastinum and pleura	1 (0.3)
C67, Malignant neoplasm of bladder	1 (0.3)
C66, Malignant neoplasm of ureter	1 (0.3)
C90, Multiple myeloma and malignant plasma cell neoplasms	1 (0.3)
D49, Neoplasms of unspecified behavior	9 (2.6)
D48, Neoplasm of uncertain behavior and of other and unspecified sites	6 (1.7)

D47, Other neoplasms of uncertain behavior or lymphoid, hematopoietic and related tissue	6 (1.7)
D21, Other benign neoplasms of connective and other soft tissue	2 (0.6)
D38, Neoplasms of uncertain behavior of middle ear and respiratory and intrathoracic organs	1 (0.3)
D35, Benign neoplasm of other and unspecified endocrine glands	1 (0.3)
D34, Benign neoplasm of thyroid gland	1 (0.3)
D25, Leiomyoma of uterus	1 (0.3)
D18, Hemangioma and lymphangioma, any site	1 (0.3)
D17, Benign lipomatous neoplasm	1 (0.3)
D12, Benign neoplasm of colon, rectum, anus and anal canal	1 (0.3)
D05, Carcinoma in situ of breast	1 (0.3)

^a All C codes were reported within 60 days before or after initiation of known sarcoma therapy; totals may add up to more than 100 percent because of multiple codes used per patient.