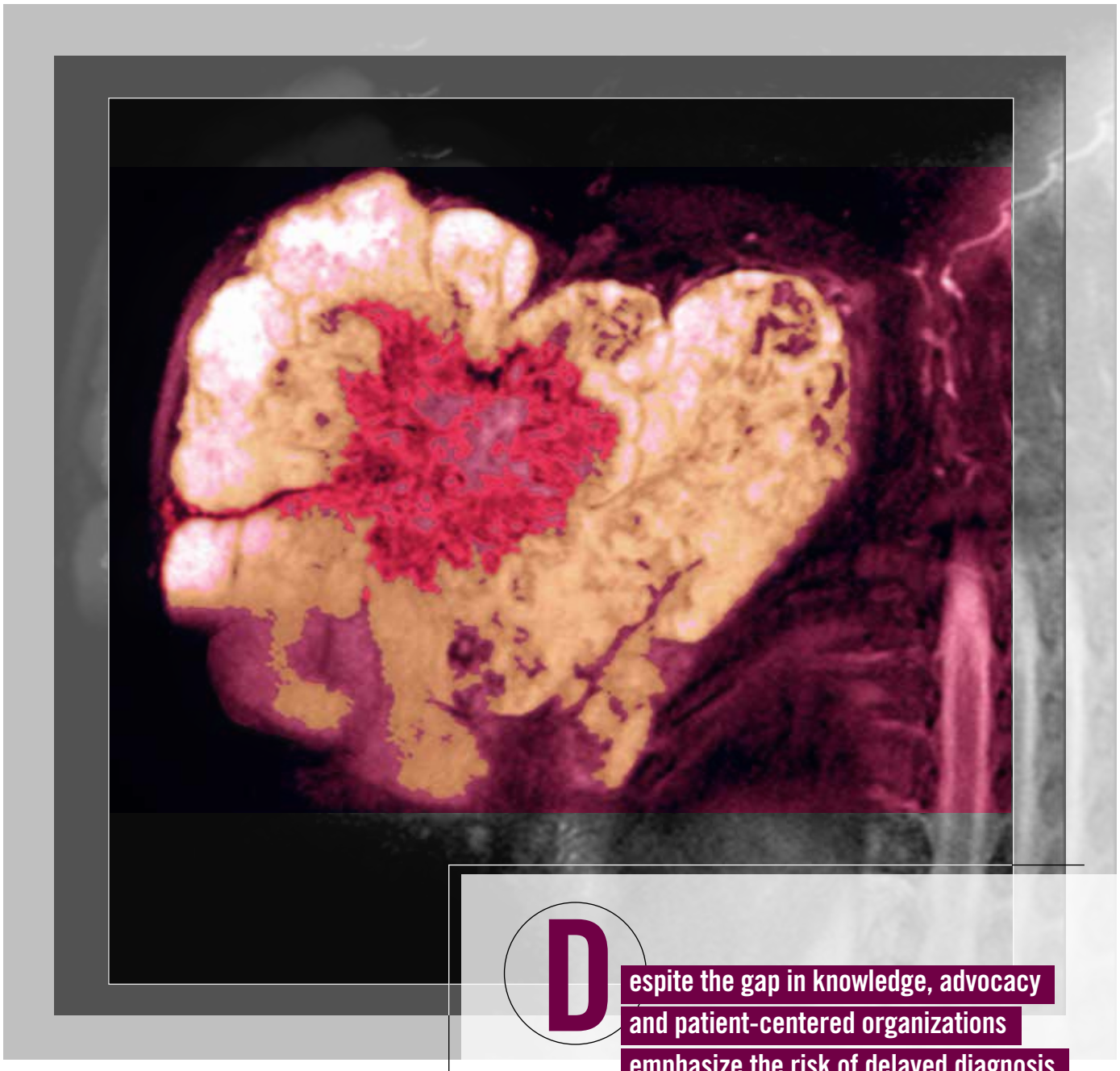

Despite the gap in knowledge, advocacy and patient-centered organizations emphasize the risk of delayed diagnosis and report high levels of stress and frustration among patients by the time an accurate diagnosis is obtained.

Health care resource utilization leading to a diagnosis of soft tissue sarcoma

ABSTRACT

Introduction The challenges of diagnosing soft tissue sarcoma are not well studied; however, the heterogeneity of its presentation would suggest that patients may experience a complex journey in the health care system prior to reaching an accurate diagnosis. This study was designed to evaluate the diagnoses, procedures, and health care resource utilization of patients with soft tissue sarcoma compared to a matched healthy control cohort.

Methods Patients in the sarcoma cohort were identified in claims data by the presence of diagnosis codes for soft tissue sarcoma. Controls were matched using exact methods on demographic, employment, and insurance variables at the date of the index sarcoma diagnosis. Health care resource utilization and diagnosis and procedure codes were compared between the cohorts during the prediagnosis period (6 months prior to the index and matched date). T test was used for continuous variables and Chi-square or Fisher's exact test was used for categorical variables.

Results A total of 7826 sarcoma patients were matched to 7826 controls on demographic, employment, and insurance variables. Diagnoses of uncertain neoplasms, limb pain, and hypertension, as well as anemia, neutropenia, thrombocytopenia, cardiac dysrhythmia, cellulitis, constipation, dehydration, diarrhea, dyspnea, edema, fatigue, gangrene, hemorrhage, nausea, pancreatitis, proteinuria, pulmonary fibrosis, rash, renal failure, vomiting, and watery eyes were significantly greater in the sarcoma cohort versus controls (all $P<.05$). The majority of health care resource utilization evaluated showed statistically higher utilization in the sarcoma cohort versus matched controls.

Conclusions Sarcoma patients had many health conditions and diagnoses that significantly differed from controls during the 6-month period prior to diagnosis. These data provide initial evidence regarding the quantity and frequency of additional health care resources used and symptoms experienced leading to the diagnosis of sarcoma.

Key words: sarcoma, diagnosis, health care resource utilization, health care economics

Lisa M. Hess, PhD,[a] Yajun E. Zhu, MS,[a] Steven Nicol, MD[a]
[a]Eli Lilly and Company, Indianapolis, IN 46285

DISCLOSURES
No funding was received or exchanged in the conceptualization, conduct, data collection, analysis, interpretation, or writing related to this study. This unfunded study was conducted by employees of Eli Lilly and Company.

INTRODUCTION

Soft tissue sarcomas (STS) are a heterogeneous group of cancerous tumors, comprised of more than 50 histological subtypes that develop from soft tissues of the body (eg, fat, muscles, nerve tissue, deep skin tissue, visceral nonepithelial tissue). Due to many factors, not limited to the heterogeneity of this set of diseases and lack of screening tests, reaching a diagnosis of STS is challenging for the general practitioner as well as for the oncologist. Sarcomas may present with nonspecific and often indolent symptomology, depending on the specific histological subtype. According to the American Cancer Society, the signs and symptoms of a sarcoma include a new or growing lump, worsening abdominal pain, blood in stool or vomit, and black stools (due to abdominal bleeding).[1] Unfortunately, these symptoms could be indicative of any number of other health conditions and are nonspecific to sarcoma.

As with many cancers, the early detection of disease when it may be completely resected could lead to a cure, whereas diagnosis when the disease is no longer amenable to surgery will impact patient survival. Among all forms of STS, early diagnosis when the patient has only localized disease is associated with an 80.8% five-year survival rate,

Due to many factors, not limited to the heterogeneity of this set of diseases and lack of screening tests, reaching a diagnosis of STS is challenging for the general practitioner as well as for the oncologists.

which decreases to 16.4% for patients whose disease has already metastasized to other parts of the body at the time of diagnosis.[2]

Previous work has evaluated the relationship between duration of symptoms that may lead to a diagnosis of sarcoma and cancer outcomes. A retrospective analysis of a cohort of adults with bone or STS found no correlation between patient recall of duration of prediagnosis symptoms and survival or metastatic disease at diagnosis.[3,4] Little other research was identified that examined the challenges of identifying a potential sarcoma. Despite the gap in knowledge, advocacy and patient-centered organizations emphasize the risk of delayed diagnosis and report high levels of stress and frustration among patients by the time an accurate diagnosis is obtained.[5] The objective of this study was to quantify the health care experience and misdiagnoses that occurred prior to a sarcoma diagnosis compared to a cohort of matched controls.

METHODS

A retrospective observational database study was conducted using detailed resource utilization and cost data from the Truven MarketScan claims database. Truven MarketScan® is a HIPAA-compliant, fully integrated patient-level database containing inpatient, outpatient, drug, lab, health risk assessment, and benefit design information from commercial and Medicare supplemental insurance plans. Additionally, the Health and Productivity Management (HPM) database, containing workplace absence, short-term disability, long-term disability, and worker's compensation data, is linked at the individual patient level. The linkage of the claims and HPM database was used for this study.

Patients were eligible for inclusion in the cohort of a sarcoma if they had at least two ICD-9 codes of 171.x on two different days between July 1, 2004, and March 30, 2014. The date of the first eligible code was considered the index date. Patients were required to have at least 6 months of health care plan enrollment prior to the first eligible ICD-9 code to allow for prediagnosis activity to be identified in the database. Patients were also required to be 18 years of age or older on the first eligible ICD-9 code date. Patients were excluded who had evidence suggesting a diagnosis of osteosarcoma, Kaposi's sarcoma, or gastrointestinal stromal tumors (treatment with methotrexate, ICD-9 codes of 176.x, 171.x, or 238.1), a history of any cancer before the eligible sarcoma ICD-9 code, or history of systemic anticancer therapy during the 6-month pre-index period. All patients meeting eligibility criteria were included in the matching algorithm to identify the control cohort.

The matched control cohort was required to have at least the same duration of follow-up at the case level as the matched sarcoma patient, could not have any evidence of any malignancy at any time in the database, nor could have received any systemic anticancer therapy at any time. Controls were randomly selected from the more than 100 million individual patient cases in the MarketScan database to be matched to the eligible sarcoma patient cohort exactly on age, geographic region of residence, health insurance plan type, gender, noncancer comorbid conditions (measured by Charlson Comorbidity Index items), and employment status. All factors were exact matched at the sarcoma cohort index diagnosis date. In the case of missing variables, patients were matched on missingness (eg, a case with missing employment status would be matched to a control with missing employment status).

The eligible time period for the index date of the possible sarcoma cohort and matched controls was between July 1, 2004, and March 30, 2014, which allowed for a minimum of 1-year follow-up through the end of the database available at the time of analysis.

All ICD-9 diagnostic and procedure codes present in the matched 6-month time period pre-index diagnosis were compared to explore factors that may be more likely to be present in the sarcoma cohort compared to matched controls. Univariate analysis was conducted for each prediagnosis variable. Analyses were

TABLE 1. Baseline demographic characteristics

Demographics		Sarcoma cohort (N=7826)	Control cohort (N=7826)	*P* value*
Gender, n (%)	Male	4069 (51.99)	4069 (51.99)	>.9999
	Female	3757 (48.01)	3757 (48.01)	
Age, years	Mean (SD)	57.8 (16.9)	57.8 (16.9)	>.9999
Duration of follow-up, days	Mean (SD)	924.5 (811.5)	1517.6 (923.8)	<.0001
Health insurance plan type, n (%)	Comprehensive	1415 (18.08)	1415 (18.08)	>.9999
	EPO	121 (1.55)	121 (1.55)	
	HMO	1170 (14.95)	1170 (14.95)	
	POS	511 (6.53)	511 (6.53)	
	PPO	4202 (53.69)	4202 (53.69)	
	POS with capitation	70 (0.89)	70 (0.89)	
	CDHP	227 (2.90)	227 (2.90)	
	HDHP	110 (1.41)	110 (1.41)	
Geographic region of residence, n (%)	Northeast	1280 (16.36)	1280 (16.36)	>.9999
	North Central	1884 (24.07)	1884 (24.07)	
	South	2855 (36.48)	2855 (36.48)	
	West	1705 (21.79)	1705 (21.79)	
	Unknown	102 (1.30)	102 (1.30)	
CCI score during pre-index period	Mean (SD)	0.4 (0.8)	0.4 (0.8)	>.9999
Number of comorbidities during pre-index period, n (%)	0 comorbidities, n (%)	5886 (75.21)	5886 (75.21)	>.9999
	1 comorbidity, n (%)	1394 (17.81)	1394 (17.81)	
	2 comorbidities, n (%)	343 (4.38)	343 (4.38)	
	3 comorbidities, n (%)	135 (1.73)	135 (1.73)	
	4 comorbidities, n (%)	35 (0.45)	35 (0.45)	
	5 comorbidities, n (%)	13 (0.17)	13 (0.17)	
	>5 comorbidities, n (%)	20 (0.26)	20 (0.26)	
Year of index date, n (%)	2004	399 (5.10)	399 (5.10)	>.9999
	2005	620 (7.92)	620 (7.92)	
	2006	438 (5.60)	438 (5.60)	
	2007	556 (7.10)	556 (7.10)	
	2008	697 (8.91)	697 (8.91)	
	2009	880 (11.24)	880 (11.24)	
	2010	1021 (13.05)	1021 (13.05)	
	2011	1141 (14.58)	1141 (14.58)	
	2012	1020 (13.03)	1020 (13.03)	
	2013	895 (11.44)	895 (11.44)	
	2014	159 (2.03)	159 (2.03)	
Employment status at baseline, n (%)	Actively employed	1145 (14.63)	1145 (14.63)	>.9999
	Not actively employed	3247 (41.49)	3247 (41.49)	
	Missing	3434 (43.88)	3434 (43.88)	

*T test was used for continuous variables and Chi-square or Fisher's exact test was used for categorical variables.
CCI=Charlson Comorbidity Index, SD=standard deviation, EPO=Exclusive Provider Organization, HMO=Health Maintenance Organization, POS=Point of Service, PPO=Preferred Provider Organization, CDHP=Consumer-Driven Health Plan, HDHP=High Deductible Health Plan

TABLE 2. Baseline clinical characteristics during 6-month pre-index period

	Sarcoma cohort (N=7826)	Control cohort (N=7826)	*P* value*
Unique diagnostic codes, mean (SD)	**12.1 (9.2)**	**7 (6)**	**<.0001**
Diagnostic codes with >10% in either cohort			
Neoplasm of uncertain behavior of skin	1186 (15.15)	0 (0.00)	**<.0001**
Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled	885 (11.31)	856 (10.94)	.46
Other and unspecified hyperlipidemia	1165 (14.89)	1082 (13.83)	.06
Benign essential hypertension	1236 (15.79)	1110 (14.18)	**.005**
Unspecified essential hypertension	1686 (21.54)	1347 (17.21)	**<.0001**
Pain in limb	824 (10.53)	407 (5.20)	**<.0001**
Localized superficial swelling, mass, or lump	1036 (13.24)	17 (0.22)	**<.0001**
Unique symptom codes, mean (SD)	2.3 (1.9)	1.9 (1.5)	**<.0001**
Symptom codes			
Anemia	662 (8.46)	353 (4.51)	**<.0001**
Leukopenia	14 (0.18)	6 (0.08)	.07
Neutropenia	12 (0.15)	4 (0.05)	**.045**
Thrombocytopenia	42 (0.54)	21 (0.27)	**.008**
Anorexia	15 (0.19)	7 (0.09)	.09
Cardiac dysrhythmias	588 (7.51)	511 (6.53)	**.016**
Cellulitis	415 (5.30)	224 (2.86)	**<.0001**
Colitis	146 (1.87)	116 (1.48)	.06
Congestive heart failure	23 (0.29)	20 (0.26)	.65
Conjunctivitis	145 (1.85)	119 (1.52)	.11
Constipation	171 (2.19)	89 (1.14)	**<.0001**

conducted using T test for continuous variables, and Chi-square or Fisher's exact test was used for categorical variables.

Number of physician visits, inpatient hospital stays, surgical procedures, and emergency room visits were compared between those in the sarcoma cohort and matched controls during the matched 6-month pre-index period. The post-index diagnosis employment status was also compared between groups using the HPM database. Comparisons between the sarcoma cohort and control cohort were made among the actively employed patients at baseline related to the proportion of patients who continued active employment, the proportion who permanently discontinued work, and the proportion who initially discontinued work and then returned to work at a later time. No adjustments were made for multiple comparisons.

RESULTS

A total of 7826 controls were each matched to patients in the sarcoma cohort. The baseline characteristics of the study cohorts are provided in **TABLE 1**. Patients with a suspected sarcoma had a mean age of 58 and were relatively balanced between male (52%) and female (48%) patients. All matched clinical and demographic variables were equivalent between groups as demonstrated in **TABLE 1**, as would be expected. The average duration of follow-up in the database was longer for the control cohort (1517.6 days, standard deviation [SD]=923.8) than for patients suspected of having sarcoma (924.5 days, SD=811.5) (*P*<.0001).

During the 6-month period before the sarcoma diagnosis (prediagnosis period), patients had significantly greater frequency of diagnoses identified than

TABLE 2. Baseline clinical characteristics during 6-month pre-index period *continued*

	Sarcoma cohort (N=7826)	Control cohort (N=7826)	*P* value*
Symptom codes *continued*			
Dehydration	86 (1.10)	39 (0.50)	**<.0001**
Diarrhea	141 (1.80)	128 (1.64)	.42
Dyspepsia/Heartburn	67 (0.86)	49 (0.63)	.09
Dyspnea	1445 (18.46)	1104 (14.11)	**<.0001**
Edema	363 (4.64)	195 (2.49)	**<.0001**
Fatigue	537 (6.86)	426 (5.44)	**.0002**
Gangrene	10 (0.13)	2 (0.03)	**.02**
Hemorrhage	784 (10.02)	507 (6.48)	**<.0001**
Hepatotoxicity	39 (0.50)	30 (0.38)	.28
Myocardial infarction	35 (0.45)	27 (0.35)	.31
Mucositis/Stomatitis	6 (0.08)	6 (0.08)	>.9999
Nausea	217 (2.77)	120 (1.53)	**<.0001**
Pancreatitis	27 (0.35)	14 (0.18)	**.04**
Peripheral sensory neuropathy	2 (0.03)	6 (0.08)	.16
Proteinuria	63 (0.81)	38 (0.49)	**.01**
Pulmonary fibrosis	64 (0.82)	17 (0.22)	**<.0001**
Rash/Desquamation	87 (1.11)	61 (0.78)	**.03**
Renal failure	109 (1.39)	64 (0.82)	**.0006**
Stevens-Johnson syndrome	0 (0.00)	1 (0.01)	**.0006**
Taste disturbance	1 (0.01)	3 (0.04)	.32
Vomiting	247 (3.16)	141 (1.80)	**<.0001**
Watery eyes	11 (0.14)	5 (0.06)	.13

*T test was used for continuous variables and Chi-square or Fisher's exact test was used for categorical variables.

controls for uncertain neoplasms, limb pain, and hypertension (all *P*<.001, **TABLE 2**). Both groups had type 2 diabetes rates higher than 10%. The symptoms patients were experiencing during the 6-month matched prediagnosis period were notable, as presented in **TABLE 2**. Most ICD codes identified in the cohorts during this period were significantly higher among those later suspected of having sarcoma, including anemia, neutropenia, thrombocytopenia, cardiac dysrhythmia, cellulitis, constipation, dehydration, diarrhea, dyspnea, edema, fatigue, gangrene, hemorrhage, nausea, pancreatitis, proteinuria, pulmonary fibrosis, rash, renal failure, vomiting, and watery eyes (all statistically significant at *P*<.05).

Similarly, the majority of health care resource utilization factors evaluated showed statistically higher health care use among patients later suspected of having sarcoma than matched controls (**TABLE 3**). Patients later suspected of having sarcoma were more likely to have surgical procedures, including an excision, resection, biopsy, or diagnostic procedure (all *P*<.0001). Blood tests were also more likely to have been performed among those diagnosed with sarcoma (41.5% vs 29.2%, *P*<.0001). Hospitalizations occurred in 15.6% of those diagnosed with sarcoma versus 7.7% among controls (*P*<.0001). Emergency room visits and physician clinic visits were also statistically significant, but the absolute rates were more modest (18.7% vs 14.6% and 94.3% vs 91.3%, respectively).

Employment status was missing for 44% of the cohort at baseline and approximately half the cohort during follow-up (**TABLE 4**). For those reporting employment,

TABLE 3. Health care resource utilization during the 6-month pre-index period (limited to those occurring in >10% of either cohort)

		Sarcoma cohort (N=7826)	Control cohort (N=7826)	*P* value
Surgical procedures				
Surgical procedure during pre-index period, n (%)	Yes	5204 (66.50)	4099 (52.38)	**<.0001**
	No	2622 (33.50)	3727 (47.62)	
Excision surgery during pre-index period, n (%)	Yes	1657 (21.17)	456 (5.83)	**<.0001**
	No	6169 (78.83)	7370 (94.17)	
Resection surgery during pre-index period, n (%)	Yes	213 (2.72)	86 (1.10)	**<.0001**
	No	7613 (97.28)	7740 (98.90)	
Diagnostic surgery during pre-index period, n (%)	Yes	2466 (31.51)	1202 (15.36)	**<.0001**
	No	5360 (68.49)	6624 (84.64)	
Other procedures				
Biopsy of skin, subcutaneous tissue and/or mucous membrane (including simple closure), unless otherwise listed; single lesion		1224 (15.64)	46 (0.59)	**<.0001**
Collection of venous blood by venipuncture		3245 (41.46)	2283 (29.17)	**<.0001**
Clinical stays/visits				
Inpatient hospitalization	Yes	1217 (15.55)	600 (7.67)	**<.0001**
	No	6609 (84.45)	7226 (92.33)	
Mean number of hospitalizations* (SD)		1.2 (0.5)	1.1 (0.4)	.10
Emergency room visits	Yes	1465 (18.72)	1141 (14.58)	**<.0001**
	No	6361 (81.28)	6685 (85.42)	
Mean number of ER visits* (SD)		1.6 (1.9)	1.5 (1.1)	**.01**
Physician clinic visits	Yes	7377 (94.26)	7142 (91.26)	**<.0001**
	No	449 (5.74)	684 (8.74)	
Mean number of physician visits* (SD)		4.7 (3.5)	3.7 (3)	**<.0001**

*Among those with 1+ hospitalization/visit
T test was used for continuous variables and Chi-square or Fisher's exact test was used for categorical variables.
SD=standard deviation; ER=emergency room

most were not employed either at baseline or during the matched follow-up period, limiting the interpretation of employment status due to the very small numbers reported. Among the eligible cohort, employment changes or retention were only reported for 960 (12.3%) patients with suspected sarcoma and 944 (12.1%) in the control group.

DISCUSSION

The symptoms experienced by patients that were recorded in claims were significantly higher across multiple categories than matched controls. However, the rates were relatively low, demonstrating the wide variability in the presentation of sarcoma. Patients had a variety of recorded problems, not limited to a lump or pain, but including hematologic, gastric, and cardiac concerns, that differed from those who had no suspected sarcoma. These factors highlight the challenges that may be facing patients who have an undetected sarcoma.

An expected finding was the difference in duration of follow-up between cohorts. This could be due to longer survival of those without a sarcoma diagnosis or due to insurance changes among those who had a sarcoma diagnosis. The absence of death data did not allow for further exploration of this finding within this study. Future research may wish to identify more comprehensive datasets to allow for the objective evaluation of the differences in time to diagnosis and stage of disease and survival, which would be the ultimate goal

TABLE 4. Post-index employment status

Post-index employment status, n (%)	Sarcoma cohort (N=7826)	Control cohort (N=7826)	*P* value*
Not actively employed at baseline, did not work during matched follow-up	3216 (41.09)	2840 (36.29)	<.0001
Actively employed at baseline, discontinued work during matched follow-up	0 (0.00)	10 (0.13)	
Actively employed at baseline, discontinued work, returned to work during follow-up	85 (1.09)	34 (0.43)	
Actively employed at baseline, continued to work during matched follow-up	875 (11.18)	900 (11.50)	
Missing	3650 (46.64)	4042 (51.65)	

*Chi-square test

in order to develop potential strategies to improve patient outcomes.

This study was limited in that the sarcoma diagnosis could not be verified in a clinical record due to the de-identified nature of the claims data used for this study. Prior work has shown that the ICD coding for sarcoma is incomplete[6,7]; therefore it is likely there are many other patients in the claims dataset who had a suspected sarcoma but who did not have a 171.x code recorded. Hence, this study is limited to a comparison of a cohort for whom the provider specified a sarcoma code in their billing records. While there are gaps in the ability to identify the entire population of sarcoma patients, the patients with ICD codes used in this study are likely true sarcoma cases. Prior work has demonstrated that the presence of these codes accurately reflects a true sarcoma diagnosis.[7] However, given the concerns with ICD coding, two sarcoma codes were required on unique days to reduce the risk of single rule-out codes or data entry error. Patients diagnosed with sarcoma demonstrate significantly greater health care resource use across variables as matched controls during the 6-month period leading to diagnosis, supporting the observations within advocacy and patient reports of the challenges faced during the process to reach an accurate diagnosis. This work may provide the initial basis for the development of strategies to more rapidly identify a potential sarcoma. Future research could also evaluate more than 6 months prior to diagnosis, to quantify the duration of time during which these differences versus controls may exist. Additionally, the cost of care may be of interest to future research to better quantify the burden of misdiagnosis on the health care system. **TSJ**

ACKNOWLEDGEMENT
The authors would like to acknowledge Yun Fang, MS, for her support in the SAS coding for the analysis of this study.

CORRESPONDING AUTHOR
Lisa M. Hess, PhD, Eli Lilly and Company. hess_lisa_m@lilly.com

REFERENCES

1. ACS. Signs and Symptoms of Soft Tissue Sarcomas. 2018. https://www.cancer.org/cancer/soft-tissue-sarcoma/detection-diagnosis-staging/signs-symptoms.html. Accessed September 27, 2018.
2. SEER. Cancer Stat Facts: Soft Tissue including Heart Cancer. National Cancer Institute Surveillance, Epidemiology, and End Results Program; 2018. https://seer.cancer.gov/statfacts/html/soft.html. Accessed February 20, 2019.
3. Rougraff BT, Davis K, Lawrence J. Does length of symptoms before diagnosis of sarcoma affect patient survival? *Clin Orthop Relat Res*. 2007;462:181-189.
4. Rougraff BT, Lawrence J, Davis K. Length of symptoms before referral: prognostic variable for high-grade soft tissue sarcoma? *Clin Orthop Relat Res*. 2012;470(3):706-711.
5. LSSI. Liddy Shriver Sarcoma Initiative. Sarcoma: A diagnosis of patience. http://sarcomahelp.org/articles/patience.html. Accessed September 20, 2018.
6. Hess LM, Zhu EY, Sugihara T, Fang Y, Collins N, Nicol S. Challenges with use of the International Classification of Disease Coding (ICD-9-CM/ICD-10-CM) for soft tissue sarcoma. *Perspect Health Inf Manage*. 2019;16 (Spring). eCollection 2019.
7. Lyu HG, Stein LA, Saadat LV, Phicil SN, Haider A, Raut CP. Assessment of the accuracy of disease coding among patients diagnosed with sarcoma. *JAMA Oncol*. 2018;4(9):1293-1295.